

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:23:46 ; Search time 146 Seconds
(without alignments)
4897.514 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAESKFNKNGKEGS.....AFVLDKNNISSKSTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2119	2 Q9AHT5_STRPN	Q9Aht5 streptococc
2	4026	100.0	2140	2 Q97RY6_STRPN	Q97ry6 streptococc
3	3793.5	94.2	2144	2 Q954M8_STRPN	Q954m8 streptococc
4	3709.5	92.1	2144	2 Q8DQ7_STRPN	Q8dqp7 streptococc
5	270.5	6.7	2649	2 Q7RAS7_PLAYO	Q7ras7 plasmodium
6	265.5	6.6	1642	2 Q8IB84_PLAF7	Q8ib84 plasmodium
7	264	6.6	3504	2 Q8IL45_PLAF7	Q8il45 plasmodium
8	260	6.5	1850	2 Q7RGP2_PLAYO	Q7rgp2 plasmodium
9	259	6.4	2757	2 Q7RRR9_PLAYO	Q7rrr9 plasmodium
10	258.5	6.4	1811	2 Q7REH9_PLAYO	Q7reh9 plasmodium
11	257.5	6.4	2661	2 Q7RMS4_PLAYO	Q7rms4 plasmodium
12	257	6.4	1389	2 Q7RPJ4_PLAYO	Q7rpj4 plasmodium
13	255	6.3	1777	2 Q8I3P4_PLAF7	Q8i3p4 plasmodium
14	253.5	6.3	1127	2 Q9YVT6_MSRPV	Q9yvt6 melanoplus
15	252.5	6.3	2227	2 Q8II21_PLAF7	Q8ii21 plasmodium
16	252.5	6.3	2723	2 Q7RQB6_PLAYO	Q7rbq6 plasmodium
17	251.5	6.2	2849	2 Q8IHY4_PLAF7	Q8ihy4 plasmodium
18	251.5	6.2	3381	2 Q8I2V4_PLAF7	Q8i2v4 plasmodium
19	251.5	6.2	3519	2 Q8IE65_PLAF7	Q8ie65 plasmodium
20	250	6.2	1474	2 Q8ILU2_PLAF7	Q8ilu2 plasmodium
21	250	6.2	1650	2 Q77J28_PLAF7	Q77j28 plasmodium
22	247.5	6.1	5767	2 Q8I525_PLAF7	Q8i525 plasmodium
23	246.5	6.1	2033	2 Q8IM18_PLAF7	Q8im18 plasmodium
24	245	6.1	2269	2 Q8ILA2_PLAF7	Q8ila2 plasmodium
25	242.5	6.0	2664	2 Q7RELO_PLAYO	Q7rel0 plasmodium
26	242.5	6.0	3317	2 Q8EWP8_MYCPE	Q8ewp8 mycoplasma
27	241.5	6.0	1033	2 Q8TBB8_PLAF7	Q8tbb8 plasmodium
28	241.5	6.0	2694	2 Q7RJJ1_PLAYO	Q7rjj1 plasmodium
29	241.5	6.0	3063	2 Q6IMC1_9APIC	Q6imc1 plasmodium
30	240.5	6.0	2740	2 Q7RFS2_PLAYO	Q7rfs2 plasmodium
31	240.5	6.0	4433	2 Q8IJ15_PLAF7	Q8ij15 plasmodium

32	240	6.0	1455	2 Q8IKG8_PLAF7	Q8ikg8 plasmodium
33	240	6.0	10061	2 Q8I3Z1_PLAF7	Q8i3z1 plasmodium
34	239.5	5.9	2569	2 Q8IBG8_PLAF7	Q8ibg8 plasmodium
35	239	5.9	2586	2 Q7PDT7_PLAYO	Q7pdt7 plasmodium
36	238	5.9	1081	2 Q8XIL2_CLOPE	Q8xil2 clostridium
37	238	5.9	1339	2 Q4YUUS_PLABE	Q4yuu5 plasmodium
38	238	5.9	3322	2 Q8IKL0_PLAF7	Q8ikl0 plasmodium
39	238	5.9	3628	2 Q968Y0_PLAF7	Q968y0 plasmodium
40	238	5.9	3704	2 Q8IKY8_PLAF7	Q8iky8 plasmodium
41	237.5	5.9	1104	2 Q7RSQ8_PLAYO	Q7rsq8 plasmodium
42	237.5	5.9	2435	2 Q4YTY1_PLABE	Q4yty1 plasmodium
43	237.5	5.9	3535	2 Q8IC29_PLAF7	Q8ic29 plasmodium
44	237	5.9	2511	2 Q8IL44_PLAF7	Q8il44 plasmodium
45	237	5.9	5229	2 Q7RTF4_PLAYO	Q7rtf4 plasmodium

ALIGNMENTS

RESULT 1
Q9AHT5_STRPN PRELIMINARY; PRT; 2119 AA.
AC Q9AHT5;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 21.
DE Serine protease (Fragment).
GN Name=prtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]_NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=N4;
RX DOI=10.1128/JAI.69.3.1593-1598.2001;
RA Witzemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a whole genome approach to identify vaccine molecules
affording protection against Streptococcus pneumoniae infection.";
RI Infect. Immun. 69:1593-1598(2001).
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
EMBL; AF291699; AAK19159.1; -; Genomic_DNA.
HSSP; P00782; 2SBT.
MEROPS; S08.064; -;
GO; GO:0009986; C:cell surface; IEA.
GO; GO:0005618; C:cell wall; IEA.
GO; GO:0016020; C:membrane; IEA.
GO; GO:0008233; F:peptidase activity; IEA.
GO; GO:0042802; F:protein self binding; IEA.
GO; GO:0004289; F:subtilase activity; IEA.
GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
GO; GO:0006508; P:proteolysis; IEA.
InterPro; IPR010435; DUF1034.
InterPro; IPR001899; Gram_pos_anchor.
InterPro; IPR003137; PA.
InterPro; IPR000209; Pept_S8_S53.
InterPro; IPR010259; Prot_inf_S8A.
InterPro; IPR001680; WD40.
Pfam; PF06280; DUF1034; 1.
Pfam; PF00746; Gram_pos_anchor; 1.
Pfam; PF02225; PA; 1.
Pfam; PF00082; peptidase S8; 1.
Pfam; PF05922; Subtilisin_N; 1.
PRINTS; PR00723; SUBTILISIN.
TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
PROSITE; PS50847; GRAM_POS_ANCHORING; 1.

```
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER
SQ SEQUENCE 2119 AA; 238228 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 5e-164;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNILGNKGKSLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 60
Db 1313 KLGEIAESKFNILGNKGKSLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 1372
Qy 61 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLEDEIKSKI 120
Db 1373 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLEDEIKSKI 1432
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVYSIEIKATKYDFHSHKTMFTFDLYANIND 180
Db 1433 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVYSIEIKATKYDFHSHKTMFTFDLYANIND 1492
Qy 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYFYVSSYGNVIELGEGDLSKN 240
Db 1493 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYFYVSSYGNVIELGEGDLSKN 1552
Qy 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILKRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1553 KPDNLTKMESGKIYSDSEKQYLLKDNIIILKRGYALKVTTYNPGKTDMLGNGVYSKEDI 1612
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMGALDGFNIIRYQVTFKQNDKGEA 360
Db 1613 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMGALDGFNIIRYQVTFKQNDKGEA 1672
Qy 361 IDKQGNLVTDSSKLVLFKGDQKDEYTGEDKNVNEAIKEDGSMFLFDTKPVNLSMDKNFYNP 420
Db 1673 IDKQGNLVTDSSKLVLFKGDQKDEYTGEDKNVNEAIKEDGSMFLFDTKPVNLSMDKNFYNP 1732
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFNWMELRVNESVDNLYLYGDLHDINTDRDFNKLNVK 480
Db 1733 SKSNKIYVRNPEFYLRGKISDKGFNWMELRVNESVDNLYLYGDLHDINTDRDFNKLNVK 1792
Qy 481 DGDIMDGMKDYKANGPPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVXPEVNID 540
Db 1793 DGDIMDGMKDYKANGPPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVXPEVNID 1852
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1853 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQHIYINGKEYTSFNDIKQIIDKTLNIK 1912
Qy 601 IVVKDFARNTTVKBEFILNKDTEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1913 IVVKDFARNTTVKBEFILNKDTEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1972
Qy 661 LEKGQVDFGWEISFEGKDKAGYVNLISKDTFIKPVFKIIEKKEEENKPTFDVSKKKDN 720
Db 1973 LEKGQVDFGWEISFEGKDKAGYVNLISKDTFIKPVFKIIEKKEEENKPTFDVSKKKDN 2032
Qy 721 PQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLDKNNISSKSTTNNPNK 773
Db 2033 PQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLDKNNISSKSTTNNPNK 2085

RESULT 2
Q97RY6_STRPN PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, entry version 22.
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641; ORFNames=SP_0641;
```

```
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtapple E.K., Khouy H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";
RL Science 293:498-506(2001).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AE005672; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSP; P00782; ZSBT.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR BioCyc; SPNE170187:SP0641-MONOMER; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR01899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inf_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF05922; Peptidase_S8; 1.
DR Pfam; PF00082; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240428 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 5.1e-164;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNILGNKGKSLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 60
Db 1334 KLGEIAESKFNILGNKGKSLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 1393
Qy 61 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLEDEIKSKI 120
Db 1394 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLEDEIKSKI 1453
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVYSIEIKATKYDFHSHKTMFTFDLYANIND 180
Db 1454 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVYSIEIKATKYDFHSHKTMFTFDLYANIND 1513
```

QY 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGVNIELGEGDLSN 240
 Db 1514 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGVNIELGEGDLSN 1573
 QY 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
 Db 1574 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633
 QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA 360
 Db 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA 1693
 QY 361 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
 Db 1694 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1753
 QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHIDNTRDFNKLNVK 480
 Db 1754 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHIDNTRDFNKLNVK 1813
 QY 481 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLTQGYDLNAKAVGVHYQFLYDNVKEPVNID 540
 Db 1814 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLTQGYDLNAKAVGVHYQFLYDNVKEPVNID 1873
 QY 541 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
 Db 1874 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 1933
 QY 601 IVVKDFARNTTVTKFELNKTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGE 660
 Db 1934 IVVKDFARNTTVTKFELNKTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGE 1993
 QY 661 LEKGQFDGWEISGPEGKADGVNLSKDTFIKPVFKIEBKKEENKPTFDVSKKDN 720
 Db 1994 LEKGQFDGWEISGPEGKADGVNLSKDTFIKPVFKIEBKKEENKPTFDVSKKDN 2053
 QY 721 POWNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSSKSTNNPNK 773
 Db 2054 POWNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSSKSTNNPNK 2106

RESULT 3
 Q9S4W8_STRPN PRELIMINARY; PRT; 2144 AA.
 AC Q9S4W8.
 DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
 DT 01-MAY-2000, sequence version 1.
 DT 07-FEB-2006, entry version 24.
 DE Cell wall-associated serine proteinase precursor PrtA.
 GN Name=prtA;
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=3.B;
 RX MEDLINE=21585565; PubMed=11728722;
 RA Bethé G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
 RA Zysk G.;
 RT "The cell wall-associated serine protease PrtA: a highly conserved
 RT virulence factor of Streptococcus pneumoniae.";
 RL FEMS Microbiol. Lett. 205:99-104(2001).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AF127143; AAD48399.1; -; Genomic_DNA.
 DR HSP; P00782; 2SBT.
 DR MEROPS; S08.064; -.
 DR GO; GO:0009986; C:cell surface; IEA.
 DR GO; GO:0005618; C:cell wall; IEA.
 DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0042802; F:protein self binding; IEA.
 DR GO; GO:0004283; F:subtilase activity; IEA.
 DR GO; GO:0004308; F:negative regulation of enzyme activity; IEA.
 DR GO; GO:0006508; F:proteolysis; IEA.
 DR InterPro; IPR010435; DUF1034.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Pept_S8_S83.
 DR InterPro; IPR010259; Prot_inh_S8A.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF06280; DUF1034; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Signal.
 FT SIGNAL 1 19 Potential.
 FT CHAIN 20 2144 cell wall-associated serine proteinase
 FT CHAIN 20 2144 PrtA.
 SQ SEQUENCE 2144 AA; 240726 MW; 2052511470741331 CRC64;

Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
 Best Local Similarity 94.7%; Pred. No. 4.7e-154;
 Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;

QY 1 KLGETAESKFNGLNGKEGSLKKTQTVGEHHHGHQENESIKKSSFTIDRNTSTIRDFENK 60
 Db 1339 KLGEIPESKFNKLVNDDSLNKETAENVENLLVDNQSIKSLFNIHKTITIRDFENK 1398

QY 61 DLKXLIKKKFRVDDFTSETGRMEYDYKYDDKGNIIAYDDGTDLLEYETEKLDKIKSKI 120
 Db 1399 DLKXLIKKKYQEDDFVN-GGTRTVRDIKYDDKGNIIAYDDGTDLLEYETEKLDKIKSKI 1457

QY 121 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIKATKYDFPHSKTMTFDLYANIND 180
 Db 1458 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIKATKYDFPHSKTMTFDLYANIND 1517

QY 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGVNIELGEGDLSN 240
 Db 1518 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGVNIELGEGDLSN 1577

QY 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
 Db 1578 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1637

QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA 360
 Db 1638 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA 1697

QY 361 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
 Db 1698 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1757

QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHIDNTRDFNKLNVK 480
 Db 1758 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHIDNTRDFNKLNVK 1817

QY 481 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLTQGYDLNAKAVGVHYQFLYDNVKEPVNID 540
 Db 1818 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLTQGYDLNAKAVGVHYQFLYDNVKEPVNID 1877

QY 541 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
 Db 1878 PKGNTSIEYADGKSVFVNINDKRNNGFDGEIQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 1937

QY 601 IVVKDFARNTTVTKFELNKTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGE 660

Db 1938 IVVKDFARNTTVKFIILNKDTGEVSELKPHRITVTIIONGKEMSTIYSEEDFILPVYKGE 1997
Qy 661 LEKGYQFDGWEISGFEGKDGAGYVINLSKDTFFIKPVFKKIEEKEENKPTFDVSKKDN 720
Db 1998 LEKGYQFDGWEISGFEGKDGAGYVINLSKDTFFIKPVFKKIEEKEENKPTFDVSKKDN 2057
Qy 721 PQVNHSQLNESHKREDLQREHSHKSDSTKDVATATVLDKNNISSKSTNNPNK 773
Db 2058 PQVNHSQLNESHKREDLQREHSHKSDSTKDVATATVLDKNNISSKSTNNPNK 2110
RESULT 4
Q8DQP7_STRR6
ID Q8DQP7_STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
AD
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocNames=spr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Essem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.,
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE008434; AAK99365.1; -; Genomic_DNA.
DR PIR; A97942; A97942.
DR HSSP; P00782; 28BT.
DR MEROPS; S08.064; -.
DR BioCyc; SPNE1313; SPR0561-MONOMER; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS0847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.

DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 2144 AA; 240438 MW; 8C1B4B1DBC503A0C CRC64;
Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 1.9e-150;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLEGIABSKPNLGNKGESLKKDITGTVVEHHHGHNEESIKESFTDIRNISTIRDPENK 60
Db 1339 KLGIEISDKLNAKSPENT-NNNAQDEKDPKESKESVEGEASLEINKTISTIRFENK 1397
Qy 61 DLAKLKKKPERVDDTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYETKLEIKSKI 120
Db 1398 DLKLLKKKPREVNDFTSETGKRIIEYDYKDDKGNIIAYDDGSALQYETKFPDEIKSKI 1457
Qy 121 YGVLSPSKOGHFEILGKISNVSKNAKVYVYNNYKSIIEIKATKYDFHSHKTTTFDLYANIND 180
Db 1458 YGVLSPSKOGHFEILGKISNVSKNAKVYVYNNYKSIIEIKATKYDHSKTTMFDLYANIND 1517
Qy 181 IVDGLAPAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYYPYSSYGNVIELSGEGLSKN 240
Db 1518 IVDGLAPAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYYPYSSYGNVIELSGEGLSKN 1577
Qy 241 KPDNLTKWESGKIYSDSEKQOYLLKONIILKRGYALKVTTYNPKTMDLSENGVYSKEDI 300
Db 1578 KPDNLTKWESGKIYSDSEKQOYLLKONIILKRGYALKVTTYNPKTMDLSENGVYSKEDI 1637
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLSALDGFNIIIRYQVTFKMDKGEA 360
Db 1638 AKIQKANPNLRVLSSETTIYADSRNVEDGRSTQAVLSALDGFNIIIRYQVTFKMDKGEA 1697
Qy 361 IDKDGNIIVTSSKLVLPFGKDDKEYTGEDKFNVAEIKEDGSMLEFIDTPKVNLSMDKNYFNP 420
Db 1698 IDKDGNIIVTSSKLVLPFGKDDKEYTGEDKFNVAEIKEDGSMLEFIDTPKVNLSMDKNYFNP 1757
Qy 421 SKSNKIYVRNPEFVLGKISDKGFNWLVRNVSVDNLIYGLDHDNTRDFNKLNVK 480
Db 1758 SKSNKIYVRNPEFVLGKISDKGFNWLVRNVSVDNLIYGLDHDNTRDFNKLNVK 1817
Qy 481 DGIIMDMGMDKYKANGFPDKVTMDGNVYLQTVGSDLNKAKAVGHYQFLYDNVKEPVNID 540
Db 1818 DGIIMDMGMDKYKANGFPDKVTMDGNVYLQTVGSDLNKAKAVGHYQFLYDNVKEPVNID 1877
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGPDGIEIQEHYVINGKEYTSFNDIKQIDKTLNIK 600
Db 1878 PKGNTSIEYADGKSVVFNINDKRNNGPDGIEIQEHYVINGKEYTSFDDIKQIDKTLNIK 1937
Qy 601 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTIIONGKEMSTIYSEEDFILPVYKGE 660
Db 1938 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTIIONGKEMSTIYSEEDFILPVYKGE 1997
Qy 661 LEKGYQFDGWEISGFEGKDGAGYVINLSKDTFFIKPVFKKIEEKEENKPTFDVSKKDN 720
Db 1998 LEKGYQFDGWEISGFEGKDGAGYVINLSKDTFFIKPVFKKIEEKEENKPTFDVSKKDN 2057
Qy 721 PQVNHSQLNESHKREDLQREHSHKSDSTKDVATATVLDKNNISSKSTNNPNK 773
Db 2058 PQVNHSQLNESHKREDLQREHSHKSDSTKDVATATVLDKNNISSKSTNNPNK 2110
RESULT 5
Q7RAS7_PLAYO
ID Q7RAS7_PLAYO PRELIMINARY; PRT; 2649 AA.
AC Q7RAS7;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN ORFNames=PY06422;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;

RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=17XNL;
 RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
 RA Carleton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
 RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 RT parasite Plasmodium yoelii yoelii";
 RL Nature 419:512-519(2002).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC -----
 CC EMBL: AABL01002172; EAA18637.1; -; Genomic DNA.
 DR GO: 0046872; F: metal ion binding; IEA.
 DR GO: 0004222; F: metalloendopeptidase activity; IEA.
 DR GO: 0009405; P: pathogenesis; IEA.
 DR GO: 0006508; P: proteolysis; IEA.
 DR InterPro: IPR011591; Botulinum.
 DR Prodom: PD001963; Botulinum; 2.
 DR Complete proteome; Hypothetical protein.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 2649 AA; 309812 MW; E7207F344643AC24 CRC64;

QY 410 NLSMDKNVFNPSKNIYVRNPEFY-----LRGLSDKGS-----FNWELRVNESVVD 457
 DB 1558 NNRKENITNLLKNENKYNVNNIFDISIRANLQMFNSNANGNESFNNEYVNNNE---H 1614
 QY 458 NYL-----IYGLDHLIDNTRDNFK---LNVRKGDIMDMGMDKYKANGFPDKVTDMDGNV 508
 DB 1615 NYTNMKNINAYSNSHLSN---NIESEHVSQNTESLQ-----NVNKKVNNKIYAFIEDV 1664
 QY 509 YLQTCGYSDLNKAKGVH-----YQFLYDNVKKPEVNIIDP 541
 DB 1665 YNRN---NLQFISLHKNDENNRVTISNGTISYNSVEKNDNSNKNYNNNSNKEYIEN 1719
 QY 542 KNTSIEYADGKSVVFNINDK-----RNNGPDGIEQHIQVINGKEYTSFNDIKQIIDK 595
 DB 1720 NHNNSIKNSD-----NVNVTYECNLKINFEFANNESTANTETENNSTNDLKNIIYE 1773
 QY 596 TLNIIK---VVKDFARNTTVKFEILN-KDTGEVSELKPHRVTVTIQNGKEMSSSTIVS-BE 650
 DB 1774 DNIIAYNNKIITENY-NDQDLKNYLYNSQKTNMNEHKPD-NMTWEKKKKGEAWIIDIKN 1831
 QY 651 DFILPVYKGELEKGYQFGWEISGEGKDGAGYVILNSKDTFIKVPFKIEBKKEENKP 710
 DB 1832 NEVLPTYTK--TEKCVLSD-----EKEAGTKKNNK-----KKSCOKGSKSNST 1872
 QY 711 -----TFDVSK-----KDNPOVN-H-----SOLNESH 733
 DB 1873 VNNNRRTYNIVKRPSSEIKNKTFFANKKNPNLHMHTKNGKLDLSLVNDISKLVNKKI 1932
 QY 734 KEDLOREHSQKSDSTKDV---ATVLDKNISSKSTNN 770
 DB 1933 KEELKGNIS-----KDTNLEITTIKKKNPISKSLNN 1966
 RESULT 6
 ID Q8IB84_PLAF7 PRELIMINARY; PRT; 1642 AA.
 AC Q8IB84;
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
 DT 01-MAR-2003, sequence version 1.
 DT 07-FEB-2006, entry version 7.
 DE Hypothetical protein MAL8P1.29.
 GN Name=MAL8P1.29;
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
 RL Quail M., Barrell B.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC EMBL: AL844507; CAD51123.1; -; Genomic DNA.
 DR Hypothetical protein.
 KW SEQUENCE 1642 AA; 194754 MW; 701F5D28B2BE8097 CRC64;
 Query Match 6.6%; Score 265.5; DB 2; Length 1642;
 Best Local Similarity 20.4%; Pred. No. 0.0059;
 Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;
 QY 4 EIAESKFNGLNGEGSLKDDTTGVEHHQNEESI---KEKSFITIDRNISTIRDENK 60
 DB 104 EINERNRNNRNNRNNNSNSFIMDEKEKTNNTIMLLKHKKRYDPPDNTS-----DEK 158
 QY 61 DLKLLIKKKPREVDFTSETGKRM-----EYDYKDDKGNIIAYDDGTDLEYTEKLEDE 115
 DB 159 NIKR-----KKINDIINKYKKMIIPSYENDSKYNSKGSFNISDD-----DE 201
 QY 116 IKSIIYGLSPSKDGHFEILGKISVSNKAKVYNNYKSIIEIKATKYDFHSTKMTFPLY 175

```
Db 202 M-----DNKSSNYNISK-----NY-STENNNDKNNPHIRSL--EKM 235
Qy 176 ANINDIVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELGEG 235
Db 236 DKINN-----EMKEY-----STKDSIDLRRVESISHSGSKE-----NNNIDKLKCH 275
Qy 236 DLSKNKPNLTKMSSGIYSDSEKQQVLLKNNILRKGYA-----275
Db 276 NLIRKNNLVNFRNK-----RKENFIFLHNYMKKKGNGKNTWINSYDIENHDEKKOK 330
Qy 276 -----LKVITYNPGK-----TDMLEGN-----GVYSKE-DIAKIQKAMPNLRALS 314
Db 331 NEILNLNDINSYNEEKCLSYNNMKENDFNKIGISKQPNVDSKERKYNKYKDHNIYNDLS 390
Qy 315 ETTIYASRNVEDGR-STQSVLMBALDGFNIIRVQVTFKQNDK-----357
Db 391 DDGKSIDRTNVDYNNSSSHNVLSNL--FNESFMLDEKIRDFKVEKVLIESLERNNS 448
Qy 358 -----GEAIDKO--GNLVTD-----SSKLVLFGKDDKEYTGEDKFNVEAIKE 397
Db 449 EIYENELKEGTHKKDELKLVLDMTIKDINEKKKKVILYHKYSE-----NYERKSISS 503
Qy 398 DGSMLFIDTPKPNLSMDKPNFSPKSKKIYVRNPEFYLRGKISDKGGFNWELRVNESVWD 457
Db 504 EN-----IDIGPTNRSFLVDGNNTKDMTHINENDR-----DKXNDN--INN--D 544
Qy 458 NYLIYDGLHDINTDFNLIKUNVKGDDIMDGMKDYKANGPDPKVTMDGNY-----YL 510
Db 545 NIIINNDNNNNNNVNVK--ESNEDIF-----YKANKLSEKIDNTVQNIIRNIKITKS 596
Qy 511 QTGYSDLNKAVGVHYQFLVDNV--KPEVNIDPKGN-----TSIEYADGKSVVFNINDKRN 565
Db 597 NSGNYNIDAN-----LSDGISNDEANLKNVNSMDYTKLYYDDDK-----GENTKN 642
Qy 566 GFDEIEQHIYINGKEYTSPNDIKQIIDTKLNKIKVVKDFARNTTVKEFILNKDTGEVS 625
Db 643 TF-----KKSAYDKNKLLENDTKLMHGNVN-----QKEWIRND--EKVLLANGINNSPK 691
Qy 626 ELKPHRVVTIQ-----NGKEMSSITVSEEDFILPVYGELEKGYQFPGWEISGPEGKKD- 680
Db 692 DLNKRNSNIDIQVNNNNNNNDTTFYDAMEYSIGNKEQYNKNVQED--HVIMGVESKN 750
Qy 681 -----AGYVNLNISKDTETI-----KPVF-- 697
Db 751 MQTWCSSNNKNYINNDRNHLKBYFDMKPGVVDNIIINIKENELVQNKKKOIFNK 810
Qy 698 ----KXIEEKKBEENK--PTFDVSKKDNPOVNH-----SQLNESH--RKEDIQRE 740
Db 811 SHGNEKIEPLKDFNNKNVKTNILKNKVNDDDIHEDVKSVINIGNEIDENRKKIKKENILNS 870
Qy 741 EHSOKSSTKD-VTATVLDKNNISSKSTTNPN 772
Db 871 INOEKDIGNKNIINTSYSKNLLSKSYAEKPH 903

RESULT 7
Q8IL45 PLAF7
ID Q8IL45 PLAF7 PRELIMINARY; PRT; 3504 AA.
AC Q8IL45;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Hypothetical protein.
GN ORFNamesPF14_0404;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12369864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
```

```
Db 2106 DNKKKFDNVLEIPKIGHNI-----LDDKETITEQVEEKSIGQDKSMENNNVST--NDG 2157
QY 640 KEMSSITVSEEPFILPVYKGELEKGYQFDGWEISGEGKKDAGYVNLK-----689
Db 2158 KOIH---IQEDI-----KGINNVNDKHSKSNLNLHIDEPKHYVEKEIKKH 2204
QY 690 ----DTPIKPVFKIEPKKEENKPT-----FDVSKKDKNPQVNHSQLNESHKREDLQREE 741
Db 2205 ETADHDIKKEFEKIEQDNKNEPSNENILVDVNAQDDK-----NISKLTNDLHDQE 2256
QY 742 HSQKSDSTKDVATVLDKKNISKS 766
Db 2257 KGTNDSVVE-----HNVSOKT 2273

RESULT 8
Q7RGP2 PLAYO
ID Q7RGP2_PLAYO PRELIMINARY; PRT; 1850 AA.
AC Q7RGP2;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN ORFNames=PY04304;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feildlyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman M.J.,
RA Carucci D.J.;
RA "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AABL01001295; EAA16146.1; -: Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1850 AA; 215317 MW; 00366A8D5CE21629 CRC64;

Query Match
Best Local Similarity 19.2%; Pred. No. 0.012;
Matches 198; Conservative 154; Mismatches 379; Indels 298; Gaps 44;

QY 6 AESKFNILNG-----KGSLLKQTTGVHHQNEEESIKEKSFITDRNISTIR-----55
Db 62 SEEKYNDQNGKGYKFNNSNTKSTNDSSYDNLADYVINE-----DNNLSFNFSVCIN 115
QY 56 ---DPEN-KOLKLLKKKF--REVDDFTSETGKRMEEYDYK---YDDKGNILAYD---D 102
Db 116 SKLDYSNILNRKSLBKPHNNNDDESDNSDEYDLYGFKNIRYKESCFNADKTKD 175
QY 103 GTDLEYETEKLDEIKSIYGVLSLP-----SKGHFPI-----134
Db 176 NAEKYSLEKYNTIDN--IGTIAPINSIGDNLKRGIKTCTNISEYNSMNNKIEVNATS 233
QY 135 ---LGKISNVSNKAVYVGNNGNYKSIETKATKYDFHSKWTTF-----172
```


Db 940 LNTQVIEK-----VKTNTVNLNEIKGLKQYDFHDFGKEENIK----- 977
 Qy 495 NGFPDPKVTMDGNYVLOTGYSDLNAKAVGHYQFLYDNVPEVNDPKGNTSIEYADGKS 554
 Db 978 --VPDEINKINEI-----KTVG--QQIDQHINKLBEIKKSGSXNDIEKGKT 1021
 Qy 555 VVP-NINDKRNNGPDGEBIQOHIYINGKEVTSFN-DIKQIDKTLNLIKIVVKDFARTTV 612
 Db 1022 DKFENVTK-----TIYNKPKKEIEEIKNI-VTKIDEKQNIYI 1059
 Qy 613 KEFTLNKDTGEVSELKPHRVT-----VTIQNGKMSSTIVSEDFILPVYKGELEKGYQ 666
 Db 1060 K---INKLLDEISEIKNNTSLKKVDINVSYGSLKLFLEQID----- 1101
 Qy 667 FDGWEISGFGKDGAGVYVNLNKDFTFKPV--FKKIEEKEEENKPTFD-----VSKKDN 720
 Db 1102 -----BEKKAGHMIK-SMEBYMDLDNKKKQEIENEMKINDIKMVKEMEA 1151
 Qy 721 PQVNHSQLNSHREKDLQREHSQKSDSTKDVATVTLDKNN--ISSKSTNNPN 772
 Db 1152 LNTSHDDDKQYH---NISKHKEEKISD-----IHKNSLKTIOBFSTESNIN 1194

RESULT 12

Q7RPJ4_PLAYO
 ID Q7RPJ4_PLAYO PRELIMINARY; PRT; 1389 AA.
 AC Q7RPJ4;
 DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
 DT 15-DEC-2003, sequence version 1.
 DT 07-FEB-2006, entry version 12.
 DE Hypothetical protein.
 GN ORFNames=PY01465;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=73239;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=17XNL;
 RA MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
 RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallow J.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
 RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 RT parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519(2002).
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NonCommercial License
 CC -----

DR EMBL; AABL01000391; EAA20802.1; -; Genomic_DNA.

DR GO; GO:0046872; F:metal ion binding; IEA.

DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.

DR GO; GO:0009405; P:pathogenesis; IEA.

DR GO; GO:0006508; P:proteolysis; IEA.

DR InterPro; IPR011591; Botulinum.

DR ProDom; PD001963; Botulinum; 1.

KW Complete proteome; Hypothetical protein.

SEQUENCE 1389 AA; 164817 MW; 21851CA630FSD9F CRC64;

Query Match 6.4%; Score 257; DB 2; Length 1389;

Best Local Similarity 20.4%; Pred. No. 0.011;

Matches 193; Conservative 139; Mismatches 288; Indels 326; Gaps 48;

Qy 17 KEGSLKQDTTGVVHHQE-----NEESIKESKSFITDRNISTIRD 56
 Db 354 KKEAFQKEKNSSEHLHEKNIFYEKEILEASKYDTVSYFNKEIKKOKSSYDNNKGWIPS 413
 Qy 57 FENKDLKLIIKKFREVDFTSETGRMBEYDYKYDDKGNIAYDDGDTLDEYETEKLDEI 116
 Db 414 EINKQRIKCNKXHFHSDT---EKVEY-----SNIIS-----DYLNQKNDK 457
 Qy 117 KSIYGVLSPKDGHFEILGKISNVKNKAKVYGN-NYKSIETKAT-----KY 163
 Db 458 KLIYGL--SKEDY---SNDKKNQDTEEVYDNDKTDIDQVNDKKNVLDISPLTHKY 510
 Qy 164 DFHSK-TMTPDLNANINDIVDGLAFAGDMELFVKDNDQKAEIKIR---MPEKIKETKSE 219
 Db 511 KYRSNLTINKINKIEKII---EINLNTDVLQKIPGMRIENVILPHK----- 556
 Qy 220 YPVSSYGVNI-----ELGEGLSKNKPNLTAKMESGKIYDSEK-----QOYL 263
 Db 557 -----YSNDILKIQIKTLSEENENIKTKE---TKNEATSYQQSDKNSQTPNSDNV 607
 Qy 264 LKONIILKGYALKV-----TTYNPGKTDMLGNGVYKED- 299
 Db 608 IKKNMIDRKTCSIKYVLIPLNNIHEIINKLILINSTYKE-----MYEKENL 657
 Qy 300 ----IAKIOKANPLRALSETTIYADSRNVEDGRSTQSVL-MSALDGFN---IIRYO--V 349
 Db 658 LPDTIYLIQ--SKSISSEIKKYMIISIKCVNGIYLYPILENVLDNFKNPMINSONNN 715
 Qy 350 FTFK-----MNDKGEAIDKD-GNLVTDSSKLVLFGKDDKEYTGEDKFENVEAIKEDGSM 402
 Db 716 STFPTVSVLYNDS---IDREIGNLJ-NTNKLJ-----NKIKIKLPEKGYI 759
 Qy 403 -----FIDTKPVNLSMOKN-----YFNPESKSNKIY 427
 Db 760 INDICFRFPDKNNVLVCVDNRKNSNGINQIGNLFIHDHGVETANYIFNLMSNFI 819
 Qy 428 VRNP-----EFLRGKISDKGFNWEIWRVNESVVDNVLIIYGDLDHN 469
 Db 820 VKSSLDLFPSPHIFPLIEQVYVDFIEKNELKLKLFIMHFLAITAICINLCFL----- 872
 Qy 470 TRDFNIKLVNKGDDIMDMGMDKYKAGFPDKVTDMDGNVYLOTGYSDLNAKAVGHYQFL 529
 Db 873 ---FNINAQI---DSLFI---KNLKAQVQIKDENDQNNFTFLYTSYEB-RVQNNKGKHPYF 923
 Qy 530 YDNVNP-----EWNIDPKGNTSIEYAD--GKSVVFNINDKRNNGDFGEIQEHIYI 578
 Db 924 LNTYSKPLIFRQSENELN-----NISHADVEGHTNLGNTSDSHRIKCEEEIKKKKEIE 976
 Qy 579 NGKEYTSFNDIKOIIDKTLNLIKIVVKDFARNTTVKEFILNKDGTGEVSELKPHRVTVTION 638
 Db 977 NG-----IK-LLKKTNDNDFIYDYARKQSI--FLYN-----AIKNDDAIKNEN 1018
 Qy 639 GKEMSSTIVSEEDF-----ILPVYKGELEK-----GYOF-----DGW 670
 Db 1019 NYDKKKQIFVTHSYTKKQFQDYVINILNRYNQLQKEKHEDYSFRREVETLQSNIRDQ 1078
 Qy 671 EISGFEGKKDAG-YVINLSKDTIKPVFKLIEKKEE-----ENKPTFDYKSKDNQVNV 724
 Db 1079 DVKSYFGKNNNIFSGGLNQ-----FDCKNEKREGIIGKSSNNKSWQNDKYTDNQNS 1131
 Qy 725 HSQLNESHREKDLQREHSQKSDSTKDVATVLDKNNISSKSTNN 770
 Db 1132 KQTFSDKKFQNIYTDENIQNLN-----NNFSTNQVNN 1165

RESULT 13

Q8I3P4_PLAF7
 ID Q8I3P4_PLAF7 PRELIMINARY; PRT; 1777 AA.
 AC Q8I3P4;
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
 DT 01-MAR-2003, sequence version 1.
 DT 07-FEB-2006, entry version 11.
 DE Hypothetical protein PFE1095w.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:49:27 ; Search time 22 Seconds
(without alignments)
792.558 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAESKFKVLNGKEGS.....ATVLKNNISSKSTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 96747 seqs, 2255637 residues

Total number of hits satisfying chosen parameters: 96747

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New.*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	4.6	1349	6 US-10-471-571A-3352	Sequence 3352, Ap
2	181.5	4.5	995	6 US-10-471-571A-3354	Sequence 3354, Ap
3	177.5	4.4	991	7 US-11-165-586-62	Sequence 62, Appl
4	177.5	4.4	1274	6 US-10-471-571A-998	Sequence 998, App
5	162.5	4.0	1169	6 US-10-527-411-20	Sequence 20, Appl
6	159	3.9	809	6 US-10-471-571A-5394	Sequence 5394, Ap
7	154	3.8	1003	6 US-10-471-571A-4826	Sequence 4826, Ap
8	153	3.8	719	6 US-10-449-902-45970	Sequence 45970, A
9	151.5	3.8	981	6 US-10-471-571A-666	Sequence 666, App
10	150	3.7	1217	6 US-10-471-571A-4942	Sequence 4942, Ap
11	149.5	3.7	872	6 US-10-527-411-145	Sequence 145, App
12	147	3.7	879	6 US-10-527-411-143	Sequence 143, App
13	147	3.7	887	6 US-10-527-411-147	Sequence 147, App
14	147	3.7	895	6 US-10-471-571A-922	Sequence 922, App
15	147	3.7	1261	6 US-10-471-571A-1304	Sequence 1304, Ap
16	147	3.7	1315	6 US-10-527-411-141	Sequence 141, App
17	147	3.7	1517	7 US-11-257-500-9	Sequence 9, Appl
18	147	3.7	1526	7 US-11-257-500-15	Sequence 15, Appl
19	146	3.6	876	6 US-10-527-411-128	Sequence 128, App
20	146	3.6	877	6 US-10-527-411-126	Sequence 126, App
21	146	3.6	877	6 US-10-527-411-130	Sequence 130, App
22	146	3.6	881	6 US-10-527-411-124	Sequence 124, App
23	146	3.6	902	6 US-10-527-411-132	Sequence 132, App
24	146	3.6	912	6 US-10-527-411-116	Sequence 116, App
25	146	3.6	914	6 US-10-527-411-120	Sequence 120, App

ALIGNMENTS

RESULT 1

US-10-471-571A-3352
; Sequence 3352, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3352
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(1349)
; OTHER INFORMATION: Ser-Asp rich
US-10-471-571A-3352

Query Match 4.6%; Score 186; DB 6; Length 1349;
Best Local Similarity 20.8%; Pred. No. 0.03;
Matches 188; Conservative 115; Mismatches 328; Indels 274; Gaps 51;
QY 23 KDTTGVHHQENE-----ESIKKSSFTIDRNISTIRDFENKOL---KK 64
Db 164 KSVNVQPTNEENKKVDAKTESTTLNVKSDAIKSNDETLDVNN-SNSNNENNADILPKS 222
QY 65 LIKKKF-----REVDD-FTSETGRMEEDYKYDDKGN-IIAYDDGTDL 106
Db 223 TAPKRLNTRMRIAIAVQPSSTEAKNVNDLITNTTLTVVD-----ADKNKIVPAQDYLSL 277
QY 107 EYETEKLDEIKS-----KIYGLVSPSKDGHFEILLGKISNVSKNAKYVYGNVVK 154
Db 278 KSOITVDDKVGSGDYFTIKYSDTVQVYG-INPE-----DIKNI---GDIKDPNNGE 324
QY 155 SIEIKATKYDFHSKTMTF---DLNANINDIVDGLAFA-----GDMRLFPVK-DN 198
Db 325 T--IATAKHDVANNLITVFTDYVDRFNSVQMGINSYIMDADTIPVSKNDVEFNVTIGN 382
QY 199 DQKAEIKIRMPKIKETKSEY-----PYVSSYGN-----VIELGEGDLSKNK- 241
Db 383 TTTTNTANIQPDYVYVNEKNSIGSAFTETVSHVGNKENPGYKQTIYVNPNSLNTAKL 442

[illegible]

```

Db 483 TTSNIENWTRNASEMVDAPK---TOEYAVTSSQVNNI-DKT---VDNEIELAPRHKD 535
Qy 573 -EQHIYINGKEYTSFNDIKQIIDKTLNLIKIVKDFARNTTVKBFILNKDTGEVSELKPHR 631
Db 536 DQTNLSVNSLKTNDVNDHVVEDSSMN-----EIEKNNAEITE----- 573
Qy 632 VTTVIQNGKMSSTIVSEBFIIPVYKGELEKGYQDFGWEISGFEGKKGADAGYVINLSKOT 691
Db 574 ---NVQNEAAESEQNEVEK-----TIENVPKQTEKYSTLSKRP 610
Qy 692 F---IKPVFKK--IEEKKEE-----ENKPTFDVSKKDNQVNHSQLNESHKEDLQREE 741
Db 611 FNVVMTSDKRRMDRRKHSKVNPELKV--QSKQAVSERMPASQATPSSRSDS-----QE 665
Qy 742 HSQSDSTKDVATVLDKNNISSKSTNN 770
Db 666 SNTNAYKTNNMTSNNVENQLIGHAETEN 694

RESULT 5
US-10-527-411-20
; Sequence 20, Application US/10527411
; Publication No. US20060110410A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130005
; CURRENT APPLICATION NUMBER: US/10/527,411
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: PCT/GB2003/003824
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 10/241,596
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 20
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-527-411-20

Query Match 4.0%; Score 162.5; DB 6; Length 1169;
Best Local Similarity 20.1%; Pred. No. 0.41;
Matches 190; Conservative 140; Mismatches 303; Indels 313; Gaps 50;

Qy 5 IAESKFNKLGNGKGSLLKDDTGVVHHQ-----ENEESIEKSSFTIDR 49
Db 319 IYKKNFKDKYFVEDSEKYSIDVESFDKLSLMFGFETNIAENVK-IKTRASYSFD- 376
Qy 50 NISTIRDFENKDLKKLIKKKFREVDFTSETGKMEBYDYKYDDKG-NITAYDDGTLEY 108
Db 377 SLPPVK-----IKNLLDNEITYTIEGENISDKDMEK-EYRGQNKAINKQAYEISKEHL 429
Qy 109 ETEKLEIKS-KIYGV-LSPSKDGHFEILGLIS---NVSKNAKYVG--NNY----- 153
Db 430 AVYKIOMCKSVKAPGICIDVDNEDLFFIADKNFSDDLKRNERYEYNTQSNYENDPFIN 489
Qy 154 -----KSEIKA-----TKYDFH----- 166
Db 490 ELIIDTDLISKIELPSENTESLTDFNVDPVVEKQPAIKKIFTDENTIFQVLYSQTFPLD 549
Qy 167 ----SKMTWFD---LYAN-----INDIVDGLAFAGDMRLFVKNDQKKAIEK 206
Db 550 IRDISLTSSFDALLFSNKVYSFFSMDYIKTANKVVEAGLPAGVWKQIV--NDPVEIANK 607
Qy 207 IRPEKIKETKSEFPYVSSYGNVIELGEGDLSKNKPNLTKWESGKI----- 253

```

```

Db 608 SNTMDKIADISLIVPI---GLALNVG-NETAKGNFENAFEIAGASILLEFIBELLIPVV 663
Qy 254 -----YDSEKQOYLLKDNILIRKG-----YALKV-----TTYNPGKTMLEGNGVYS 296
Db 664 GAFLLSEYIDNKNKIITIDNALTNRNEKWSMDYGLIVAQWLSTVNTQFTYIKEG----- 718
Qy 297 KEDIAKIQKA-NPNLRALSETTIYADSRNVDEGRSTQCVLSMALSALDGFNIIRYQVFTFKMN 355
Db 719 -----MYKALNYQAQALEEIIKYRYNIYSEKEKSNINI-----DFNDI-----NSKLN 761
Qy 356 DK-GEAIDKDGNLVT-----DSSKLVLPFGKD 380
Db 762 EGINQALDINNINPFINGCSVYLMKKVPIPLAVEKLLDFDNTLKKNLNYIDENKLYLIGSA 821
Qy 381 DKEYTGEDKFNVEAIKEDGSMLEFIDIKPVNLSMDKNYFNPSKSNKI----- 426
Db 822 EYEKSKVNYKLTIMPFDLSIYTNDIIEFMF---NKNYSEILNLIILNRYKDNLLIDL 878
Qy 427 --YVRNPEFYLRGKISDKGFF-----NWEELRVNESVVDNYLIYGDHIDTRDFNKL- 477
Db 879 SGYGAKEVYVGVDELNDKNQFKLTSSANSKIRVTQ---NQNIIFNSVFLDFSVSFWIRIP 935
Qy 478 NVKXGDDIMDMGMDY-----KANGFEDKVTMDGNVLOTGYSDLANAKAVGVHYQFLYDN 532
Db 936 KYKNDGIQYIHNEYTIINCWKNSGKIS-IRGNRIWT-LIDINGKTSVFFEY---N 990
Qy 533 VKPEVNIDPKGNTSIEYADGKSVVFVINDKRNNGFDEIQEIOHQHYINGKEYTSFNDIKQI 592
Db 991 IREDIS-----EVIN-RWPFVITNNLNA-----KIYINGK-LESNTDIKDI 1031
Qy 593 IDKTLNLIKVVK--DFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSE 649
Db 1032 REVIANGEIIFKLDGDIRT---QFIWMK-----YFSIFN-TELSQSNIIE 1073
Qy 650 EDFILPVYKGELEKGYQDFGW-----BISGFEGKKDAGYVINLSKDTFIKPVFKKIE 701
Db 1074 R-----YKIQSYSEVLKDFWGNPLMYNKEYYFNAGNKNKY-IKLKDS---PVGELIT 1123
Qy 702 EKKEENK-----PTFDVSKKKONPQVNHSQLNESHKED 736
Db 1124 RSKYNQNSKYINVRDLVIGEKFIIRKSNQSIN---DDIVRKED 1165

RESULT 6
US-10-471-571A-5394
; Sequence 5394, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5842
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 5394
; LENGTH: 809
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(809)
; OTHER INFORMATION: hyaluronate lyase precursor
US-10-471-571A-5394

Query Match 3.9%; Score 159; DB 6; Length 809;
Best Local Similarity 20.6%; Pred. No. 0.37;
Matches 162; Conservative 114; Mismatches 301; Indels 208; Gaps 42;
Qy 69 KFEVDDF---TSETGKRMEEYDYKDDKGNIIAYDDGTDLVEYTEKLEIKSKIYGVLS 125

```

Db 30 EFRSVDKHQIAVADTNVQPDYK-----LRNTWLDVNYGYDKYDE-----NN 72
QY 126 PSKDGHEIFLGK-ISNVSKNAKYVYGNVYKSIKATKYDFHSKTMTPDLYANINDIVG 184
Db 73 PDMKKFDATEKATNLLKEMKTESGRKY--LWSGAETLETNSHMT-RTYRNIKIAEA 129
QY 185 LAFAGDMRLFVKDNDQKKAIEKIRMPKIKETKSEYPYVSSYG-----NVIELGEG---D 236
Db 130 MR---NPKTTLNTDENK-----KVKDA-LEWLHNAYGKBPDKVKELSENFTKT 176
QY 237 LSKNKPDLNLTWESGKIYSDSEKQOYLKDNIL-----RKGYALKVTTYNP----- 283
Db 177 TGKNTNLNWDVEIGTPKS-----LNTLLILNDQFSNEEKKKFTAPIKTFAPDSDK 228
QY 284 -----GKTDMLEGNVYSKEDIKIOKANPNLRALSETTIYADSRNVEDGRSTOSVLSA 338
Db 229 ILSVSGKAEALAGNLV---DISVKV-----LLECIIEB-----KMMKKS 267
QY 339 LDGFNIIRYQVTFKNDKGEAIDKDGNLVTDSSKLVLFGK-----DKSYTGEDKFN 391
Db 268 IDSFN---KVFTY-VQDSATGKERNG-----FYKDGSIYDHQDVPYTG--AYG 309
QY 392 VEAIKEDGSM-LFIDTKPNVLSMDKNYFNPSKSNKIYVRN-----PEFYLRGKISDKG 443
Db 310 VVLEGISQMPMIKETP-----FNDKTQNDTTLKSWIDDGFMPLIY-KGEMMDLS 359
QY 444 GPNWELRVNE-----SVVDNVLIIYDGLHIDNTRDFNKLNVKDGIDMDGMDKYKANG 496
Db 360 RGRALSRENETHSHASATVWKSLLRLSDAMDSTK-AKYKKIYKSVESD-----SSYKQND 415
QY 497 FPDKVTMDGNVYLTQGYSDLNKAVGVHYQFLYDNVKNPEVNIDPKGNTSIEYADGKSVV 556
Db 416 YLNSYSDIDKMSLMT---DNSISKNGLTQQLKIYNDMDRVYH---NKDLOFAFGLSMT 469
QY 557 FNINDKRNGPGEQOEQHIYNGKEYTSFNDIKQIIDLKTLNIIKIVVKDFARNTTVKEPI 616
Db 470 -SKNVARYESINGENLKGWHTGAGMSYLYNSDVGVHYD-NFVWVTADMKRLSGTTLIDNEI 527
QY 617 LNKDTGEVSELPHRVTVTIQNG-----KEMSTIYSEEDFIL---PVYK 658
Db 528 L-KDTDKSSK-----TFVCGTKVDHDSAGHMDFENQDKTLTAKKSYFILNDKIYVL 580
QY 659 GELEKGYQPDGWEISGFEKKDAGVYNLSKDTFIKPVFKKIEKEEENKPTF---DV 714
Db 581 GTGIKSTDSKPNVTTIENRKANGTYLYTDD-----KQTTNSDNQENNSVFLESTDT 632
QY 715 SKKNDPNQVNSQL---NESHK---KEDLQREHSQKSDSTKDVATVLDKNNISSKSTT 768
Db 633 KKNIGYHFLNKPKITVKKESHTGKWKINKSOKTKTDEYYEVT-----QKHS 681
QY 769 NNPKN 773
Db 682 NSDNK 686

RESULT 7
US-10-471-571A-4826
; Sequence 4826, Application US/10471571A
; Publication NO. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 4826
; LENGTH: 1003
; TYPE: PRN

; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(1003)
; OTHER INFORMATION: chromosome segregation SMC
US-10-471-571A-4826

Query Match 3.8%; Score 154; DB 6; Length 1003;
Best Local Similarity 18.4%; Pred. No. 0.9;
Matches 157; Conservative 146; Mismatches 335; Indels 214; Gaps 39;

QY 5 IAESFKNLGNKESGLKKOT--TGVEHHHQBNEBSIK-----EKSSFTIDRN--ISTIR 55
Db 49 LGEQSAKSL---RGSKMEDIIFSGAEHRKAQYAEVOLRLDNHSHKSLSDVENEIVITRR 104
QY 56 DFNKDLKKLI---KKKPREVDFTSETCKRMEY-----DYKYDDKGNIIAY 100
Db 105 LYRSESEYYINNDRARLKDIADFLDGLGLGKEAYSIISQGRVDEILNAPKIDRRQIIEE 164
QY 101 DDGTOLEYE-----TEKLDEIKSIYGVLSPSKDGHEILGKISNVSKNAKYVYGNVYK 154
Db 165 SAGV-LKYKKRKAESLKLQDQEDNLTRV-----EDILDLEGRVPELKEAAI--AKEYK 217
QY 155 SIEIKATKYDFHSKTMTPDLYANINDIVDGLAFAGDMRLFVKDNDQKKAIE-KIRMPK 213
Db 218 TLHQMKHSDIVVTVHDIDQVTDNDRQL-----DQRLNDLQGGQANKEADKQRLSQOI 270
QY 214 KETKSE-----YPYVSSYGNVIELGEGDLKSNKPNLTWESGKIYSDSEKQO 261
Db 271 QOYKGRHQLDNVDSLNYQLVKA-TEAFKYTGQNLVLEERKKNQSQSTNARYEEQENL 329
QY 262 YLLKNDIILRGYALKVTTYNPCKTDMLEGNVYSKEDIKIOKANPNLRALSETTIYAD 321
Db 330 MELLENI-----SNEISEAQDTY-KSLSKOKELNAVIRELEEQLYVSD 372
QY 322 ---SRNVEDGRSTQSVLMS-ALDGFNIIRYQVTFKMDN-KGEAIDKGNLVTDSKLYL 376
Db 373 EAHDEKLEIKNEYTYTLMSEQSDVNDIRFLKHTTEENEAKKSRLD-----SRLVE 423
QY 377 FGKODKEYTGEDKFNVEAIKEDGSMFLFDTKPNVLSMDKNYFNPSKSNKIYVRNPEFYLR 436
Db 424 VPEQLKDIOGQIKTKTKKYQOTNKELSADVKEIK-NIEKDLTDTTKAQNVEY--EEKLYQA 480
QY 437 GKISDKGFNNELNRVNESVDNLYLYDGLHIDNTRDFNIK-LNVKDGIDMDGMDKYKAN 495
Db 481 YRYTEKMTRIDSLATQEEYTYFNGVKHILKAKNKKELKGIYGAVAEIID-----531
QY 496 GPPDKVTMDGNVYLTQGYSDLNKAKAVGVHYQFLYDNVKNPEVNIDPKGNTSIEYAD---551
Db 532 -VPSKLTQA-----IET-----ALGASLQHV-----VDSEKDGQRAIOFLKERNL 571
QY 552 GKSUVFNIN-----DKRNNGPDGEQOEQHIYNGKEYTSFNDIKQIIDLKTL 597
Db 572 GRATFLPLNVIQSRVVATDIKSIKANGP-----ISIASEAVKVAPEYQNTIGNLL 623
QY 598 NIKIWKDFARNTTVKEFILNKDTGEVSELKPHRV-TVTIO-----NGKEMSS 644
Db 624 GNTIIVDHL-----KHANELARAIKYRTRIVTLEGDIYVPGSGMTGGARKSK 671
QY 645 TIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGVYNLSKDTFIKPVFKKIEKK 704
Db 672 SILSKQD-ELTTRMRHQLEDYLR---QTESFEQQ-----FELKTKS 708
QY 705 EEENKPTFDVSKKD--NPQVNSQL-----NESHKEDLQREHSQKSDSTKDVAT 755
Db 709 DQLSELYFEKSKNTLKEQVHHFEMELDRLTQETQIKNDHPEFEFEKNDGYTSDKS---766
QY 756 VLDKNNISSKST 767
Db 767 ---RQTLSEKET 775

Db 583 DEL---QSTVE-----||||| 594
QY 507 NVYLOTGYSDLNKAVGVHYQFLYDNVKEPVNIDPKGNTSIEYADGKSVFVNIDKRNNG 566
Db 595 ELYLSKSN-----DUIVSISTWANKALDQHSIDLNQORQOLVOELDTFFNH 644
QY 567 FGEEIQEQHYINGKEYTSFNDKIIDKTLNIIKIVVKDFARNTTVKEFILNKDTGEVSE 626
Db 645 ABAAVTKSQPVYFN--KLSLFHDVQWLKS-----AEDTNEKWRINAENTKLVN 691
QY 627 LKPHRVTVTIOGKEMSS-----TIVSEEDFILPVYKGELEKYQPDGWEISGFEG----- 677
Db 692 ELNHLNAQLEANNKEITALDFINVGTEEDFY-----QHEDYQTYTNSLRFNDLTKY 745
QY 678 KKDAGVYVLSKDTFKIPFKKIEE-----KKEENKPTDPVSKKD--NPQVNHSQLN 729
Db 746 LENQVSYVLSLSSSEKTTAQLLEEDHLLATQVDEYNEQYLEMAQVSDLSAQINMETD 805
QY 730 ESHRKEDLQREHS---QKSDSTKD 751
Db 806 TT--LANLRHEYHSLKNQNDIAKD 828

RESULT 10

US-10-471-571A-4942
; Sequence 4942, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: Seqmin99, version 1.03
; SEQ ID NO 4942
; LENGTH: 1217
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1217)
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-4942

Query Match 3.7%; Score 150; DB 6; Length 1217;
Best Local Similarity 20.4%; Pred. No. 1.9;
Matches 180; Conservative 135; Mismatches 320; Indels 248; Gaps 46;

QY 39 IKEKSSFTIDRNISTRD--FENKDLKKLKKK-----PREVDDFTSETGR----- 83
Db 289 IKEANEMVD-----ALEDAKQYKKYKSLIDKVSDFSRADDLKADMQQLAPRVKYLA 344
QY 84 -----MBEYDYKDDKGNIIAYDDGTDLLEYETEKLDEIKSIYGLVSPSKDG-----HF 132
Db 345 RIVKDVMSFNRKESK-NILDFSD-----YEHFAL-QILTNEGD--SPSEIAESYRQHF 395
QY 133 E-----ILGKISNVSKNAKYVYGNVNNYKSIETKATYDPH--SKTWTFDLYANI 178
Db 396 QEILVDEYQDTRVQGEKILSCITKIDGHNGLFMVGDVKQSIYKPRQADPSLFIKRYQRF 455
QY 179 NDIVDGLAFAGDMRLFVKDNDQKAEI-----KIRMPKIKETKSEYFPVSVSGNVI 230
Db 456 TIDGGTG-----RRIDLQNFRRSEVLSTTNIIFKHMWDEQGVGEVKYDEAAQLYYGAPY 511
QY 231 ELGEGDLSKNKPNL--TKMESKIIYSD---SEKQOYLL-----KDNILRKGYALKVTTYN 282
Db 512 D-----ESDHPVNLKVLVEADQEHSLTGSEQEAHFIVEQVKDILEKQKYDMKTGYSR 565
QY 283 PGKTDMLGNGVYSKEDIATQKANPNLRALSETTIYAD---SRNVEDGRSTQSVL----- 335

Db 566 SA-----TKDIVILERSFGQARNLQAFKNEIDIPPHVNSREGYFPEOTEVRVLV 614
QY 336 --MSALDG-----FNIRYQVFTFKMNDKCEADKDGNLVTDSSKLVLFKDKDKKEYTG 386
Db 615 SFLRAIDNPLQDIYLVGLMRSYVIQFKEDELAQI-----RILSPNDDDYFYQS 661
QY 387 -EDKFNVEAIREKDGSMFLDTKPVNLSMDKNYFNPSKSNKIYVRNPEF-----YLRG 437
Db 662 IWYINDEA---ADAILVDKLMFSLDIQSYYQYQSKDHPVYQLDKFNDHVVHVIQYFSG 717
QY 438 KISDKGG-----FNWELRV-----NESVV---DN----- 458
Db 718 LIGGRGRANLYGLENKAEIENFSFRGLYQFIRFIDELIERGKDFGEENVVGVNDNVVR 777
QY 459 -----YLIYGDHLID-NTRDFN--IKLVNKGDDIMDGMKDYKANGPDKVT 502
Db 778 MWTIHSSKGLFPFFVIYSGLSKDFNKRDLKQPVILNQOFGLGMDYFDVD-KEMAF----- 832
QY 503 DMDGNVYLQTYGSDLNKAVGVHY-----QFLYDNVVKPEVNIDPKGNTSIEYADGK 553
Db 833 SLASVAYRAVAEKELVSEEMRLVYVALTRAEQLYLIGRVKNDKSILLEQLSI---SGE 889
QY 554 SVVFNIDKRNNGFD--GEIOEQHYIINGKEYTSF-NDIKQIIDKT---LNKIVVKDPA 607
Db 890 HIAVNERLTSPNPFHLIYSILSKHQASIPDDLKFEKDIAQIEDSSRPNNVISIV---YF 946
QY 608 RNTTVKEFILANKDTGEVSELKPHRVTVVTIOGKE--MSSTIVSEEDFILPVYK----- 658
Db 947 EDVSTETILDNDEYRSVNOLE-----TWQNGEDVKAQIKHOLDYRYPVNDTKPKSKQ 1000
QY 659 --GELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVS- 715
Db 1001 SVSELKROYE---TRESGTSYRVRQYRIGFS--TYERPKEFSEOGKKAKEIGTLMHTV 1055
QY 716 -----KKKDNPPQVNHSQLNESHRKEDLQREHSQKSDSTKDV 752
Db 1056 MQLHPFKKERISEV-----ELHQYIDGLIDKHIIEADAKKDI 1092

RESULT 11

US-10-527-411-145
; Sequence 145, Application US/10527411
; Publication No. US20060110410A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Keith Alan
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130005
; CURRENT APPLICATION NUMBER: US/10/527,411
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: PCT/GB2003/003824
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 10/241,596
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 145
; LENGTH: 872
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-527-411-145

Query Match 3.7%; Score 149.5; DB 6; Length 872;
Best Local Similarity 20.4%; Pred. No. 1.3;
Matches 131; Conservative 107; Mismatches 230; Indels 175; Gaps 36;
QY 79 ETGRMEEYD-----YKYDDKGNIIAYDDGTD--LEYETEKLDEIKSIYGLVSPS 127

Db	51	EFGTPEDFNPSSLIEGASEYDP--NYLRDTSOKRFLQTMVKLFNRKNNVAGE----	105
Qy	128	KDGHPEILGKISVAGSKARVYGGNNYKSIETIKATKYDFHSKTTFFOLY-ANINDIVIDGLA	186
Db	106	-----ALLDKII---NAIPLGNSYSLLD---KFDTSNSVSNFLLEQDPGSATTKSA	152
Qy	187	PAGDMRLFVKNDQKKAIK--TRMEPIKETHSEYP-----YVSSVG	227
Db	153	MUTNIIIFGPGVULNKNRVRGILURV-----DNKNYFPCRDGFGSIMQAFCEYPTFD	207
Qy	228	NVIELGEGDSLKNKPDNLTQWESGKIYSDSEKQYLLKDNII-ILRKGYALKVTTYN--PG	284
Db	208	NWIE-----NITSLTIKSKYFQDPALLAHLELHVLHGLYGNQVSSHEIIPS	255
Qy	285	KYDMLE-----GNGVYSK-----EDIAKIQKAN-P	308
Db	256	KQEIYMQHTYPIASABELFTFGQDANLISIDIKNDLYEKTLDNYKAIANKLSQVTSNDP	315
Qy	309	NLRALSETTIYADSRNVEDGRSTQSVL-----MSALDGNII---RYQVFT----	351
Db	316	NIDIDSQKIYQOKYQPKDSNGQYIVNEDKFQILYNSIMVGTEIELGKFKNIKRLSY	375
Qy	352	FQWNDKGEAIDK--DCNLVTUSSKLVLFGDDK--EYTGED-KFNVEAIKE-DGSMLEFDT	406
Db	376	FSMNHDPVKPIPNLLDDTTIYNDTEGFNIESDKLSEYKGQNMRYNTWAFRNVDSGLV--S	433
Qy	407	KPVNL--SMDKNYFNPKSN-----KIYVNPFEYLRGKISDKGFGFNWELRVNESVDN	458
Db	434	KLIGLCKSEELKYDDDDKRWGSSRICTIKNEDLTF---IAEKNFS-----EEFPQDE	485
Qy	459	YLIYGDLDHIDNTRDNFIKLNVK-DGDIMWGKQDYKANGSPDKVTDMDGNVYLQCYSDL	517
Db	486	IVSY-----NTKNKPLNFNSLQKIIYDVNLOS-KITLPNDRTPVTPTGIPVAPKYSN	538
Qy	518	NKAVGVH-----YQFLYNNVAKPEV-----NIDPK--GNTSIYADGKSVNENIN	560
Db	539	AASTTEIHNIIDNTIYQVLYAQKSPFTLQRTWTNSVDDALINSTKI-YSYFPSVI----	593
Qy	561	DKRNGFGGEIQEQHIYINGKEYTSFNDDIKQIDKTLNIIKIV	603
Db	594	SKVNOCAQGIPLQWVRDIIIDFNFNESSQKTTIDKISDVSTIV	636

RESULT 12

```

US-10-527-411-143
; Sequence 143, Application US/10527411
; Publication No. US20060110410A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stuncombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130005
; CURRENT APPLICATION NUMBER: US/10/527,411
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: PCT/GB2003/003824
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 10/241,596
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 143
; LENGTH: 879
; TYPE: PRP
; ORGANISM: Clostridium tetani
US-10-527-411-143

```

Query Match

3.7%: Score 147: DB 6: Length 879:

Best Local Similarity 20.0%; Pred.No. 1.7;		
Matches 130;	Conservative 110;	Mismatches 228; Indels 192; Gaps 36;
QY	79	ETGKMEEDY-----YKYDDKGNIIAYDDGTD--LEYETEKLEDBEIKSIYGVLSPS 127
Db	51	EGTGPEDPNPSSILEGASEYDP-NYLRTDSDKDRFLQTMVKLFNRKKNVAGE----- 105
QY	128	KDGHFEILGKISVSKNAKVYGNNYKSIIEIKATKYDFHSKTMFTDY-ANINDIVDGLA 186
Db	106	-----ALLDKII---NAIPYLGNSYSLD---KFDTSNSVSFNLLLEQDPSGATTKSA 152
QY	187	PAGDMWLFVKMDQKKAIK---IRAPKIKETKSEYP-----YVSSYG 227
Db	153	MLTNLIIFGPGVPLNKNVIRGLRV-----DNKNYFPCRDGFGSIMQAFCEYVTFD 207
QY	228	NVIELGEGDLSKNKPNLTKMESGKIYSDSEKQOYLLKDN-IIRKGVALKVTYN--PG 284
Db	208	NWIE-----NITSLTIKSKYFQDPALLMHELHVLHGLYGMQVSHSEIIPS 255
QY	285	KTDMLE-----GNGVYSK-----EDIATKOKAN-P 308
Db	256	KQEIIYMQHTPIASAEELFFGQDANLISIDIKNDLYEKLTDYKAITANKLSQVTSNDP 315
QY	309	NLRALSETTIYADSRNVEDGHSQTQSVL-----MSALDGFNII---RYQVFT----- 351
Db	316	NIDIDSYKQIYQOKYQFQDKNGQYIVNEDKFQILYNSIMYGFEIELGKKFNITKRLSY 375
QY	352	FKMNDKGEAIDK--DGNLWTPSSKLVLPFGDKD-EYTGED-KFNVEAIKE-DGSMLFIDT 406
Db	376	FSMNHDPVKIPLLDDTIYNTEGFIENISKOLKSEYKQGNRWVNTAFNRVDSGLV--S 433
QY	407	KPNVL-----SMDKNYFNPSKSNK-----IYVRNPEFYLRGKISDKGGFNWELRV 451
Db	434	KLIGLCKKIIPTNIRENLXNRTASITDLGELCLCIKNEDLTF---IAEKNFSF----- 485
QY	452	NESVVDNYLIYGDLIHIDTRDENIKLVK-DGIDMDWGMKDYKANGPDKVTDMDGNVYL 510
Db	486	BEFFQDEIVSY-----NTKNKPLNFNYSLKIIIVDYNLOS-KITLPNDRTPPTVKIGIPY 538
QY	511	QTGYSDLNAKAVGVH-----YQFLYDNVKPEV-----NIDPK--GNTSEIYADGK 553
Db	539	APYEYKSNAASTIEINIDNTIYQVLYAQSFITLQRTWMTNSVDDALINSTKL-YSYFP 597
QY	554	SVVFNINDKRNNGFGEIQEQHIYINGKEYTSPNDIKQIIDKTINIKIV 603
Db	598	SVI-----SKVNOAGCILLFLOWVRDIIDFTNBSOSKTIKIDKIDSVTIV 643

RESULT 13

```

US-10-527-411-147
; Sequence 147, Application US/10527411
; Publication No. US20060110410A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130005
; CURRENT APPLICATION NUMBER: US/10/527,411
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: PCT/GB2003/003024
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 10/241,596
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 147
; LENGTH: 887
; TYPE: PRT

```

```

; ORGANISM: Clostridium tetani
US-10-527-411-147

Query Match      3.7%; Score 147; DB 6; Length 887;
Best Local Similarity 20.0%; Pred. No. 1.8; Mismatches 182; Gaps 36;
Matches 130; Conservative 110;

QY 79 EFGKRMEEYD-----YKDDKGNIIAYDDGTD--LEYETEKLDEIKSIYGVLSPS 127
Db 51 EFGKPEDEFNPSSLEIAGASEYDYP-NYLRTSDXDRFLQTVKVLFNRIKNNVAGE--- 105

QY 128 KUGHPEILKISNVSKNAKVYGNVYKSIIEIKATKYDFHSHKWTWFDLY-ANINDIVDGLA 186
Db 106 -----ALLDKII-----NAIPYLGNSYLLD---KPDNTNSVSNFLLQDPDSGATTKSA 152

QY 187 FAGDMELFVKDNDQKAEIK--IRMPEKIKETKSEYP-----YVSSVG 227
Db 153 MTLNLIIFGPGVPLNKNEVRGIVLRV-----DNKNYFRCRDFGSGIMQWAFCEYVPTFD 207

QY 228 NVIELGEGDLSKNKPNLTKMBSGKIYSDSEKQQVILLKDNII-ILRKGVALKVTYTN--PG 284
Db 208 NVIE-----NITSLTIGSKYFQDPALLMHLELIHVLHCLGMQVSSHEIIPS 255

QY 285 KTDMLB-----GNGVYSK-----EDIAKIQKAN-P 308
Db 256 KQEIYMQHTYPIASBELFTFGQDANLISIDIKNDLYEKTLDYKAIANKLSQVTSQNDP 315

QY 309 NLRALSETTIYADSRNVEDGRSTQSVL-----MSALDGFNII---RYQVET----- 351
Db 316 NIDIDSYKQIYQOYQFQKDSNGQYIVNEDKQFQILYNSIMYGFTEIELGKFNKIRLSY 375

QY 352 FRMKDKGEAIDK--DGNLVTDSKLVFGKDDK-EYTGED-KFNVEAIE-KDGMFLDIT 406
Db 376 FSNMHPVKIPNLLDDTIYNDTEGNIIESKDLKSEYKQNMVNTNAPFNVDSGLV--S 433

QY 407 KPNVL-----SMDKNVFNPSKNNK-----IYVRNPEYLRGKISDKGGFWELRV 451
Db 434 KLIGLCKKIIPPTNIRENLNRYNTASITDLGGELCIIKKNEDLTF---IAEKNSFS----- 485

QY 452 NESVDNLYLYDGLHDNTRFNKLVNK-DGDMWGMKDYKANGFPDKVTDMDGNVYL 510
Db 486 BEFPQDEIVSY-----NTKNKPLNFNYSKDKIIVDYNLQS--KITLPNDRTTPVTKGIPY 538

QY 511 QYGYSDLNAAKAVGVH-----YQFLYDNVKEPV-----NIDPK-GNTSIEYADGK 553
Db 539 APEYKSNAASTIEHNIDNTIYQIYLAQKSPITLQRTWNSVDALINSTKI-YSYFP 597

QY 554 SVFVNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNKKIIV 603
Db 598 SVI-----SKVNOGAQGILFLQWVRDIIIDFTNESSQKTTIDKISDVSTIV 643

```

```

RESULT 14
US-10-471-571A-922
; Sequence 922, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 922
; LENGTH: 895
; TYPE: PRP
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1) .. (895)

```

```

; OTHER INFORMATION: hypothetical protein
US-10-471-571A-922

Query Match      3.7%; Score 147; DB 6; Length 895;
Best Local Similarity 18.5%; Pred. No. 1.8; Mismatches 237; Gaps 28;
Matches 129; Conservative 87;

QY 77 TSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVEYETEKLDEIKSIYGVLSPSKDGHFELIG 136
Db 392 TASTWKKEFEVGEKGLPELVLSYSDSKDAY-----TRP 426

QY 137 KISNVSKNAKVY-----YGNVYKSIIEIKATKYDFHSHKWTWFDLYANINDIVDGLAFAGDMR 192
Db 427 PVSNGTREVKIVSSIEYGENIHE-----DYDYLTLWFAQPIITNPNDDYVDEETY-NLQK 479

QY 193 LFKVNDNDKKAIEIKIRMEPEKIKETKSEYPYVSSYGNVIELGEGDLSKNKPNLTKMBESGK 252
Db 480 LLAPYHKAKTLERQVYIELEKLOEKLPE-KYKAEYKCKLDQTRVELADQVKSATVEFEN-- 536

QY 253 IYSDSEKQOYLLKDNIIILRKGVALKVTYTNPGKTDMLGNGVYSKEDIKQKPNLRA 312
Db 537 -----VPTNDQLTDLQEAHFV----- 553

QY 313 LSETTIYADSRNVEDGRSTQSVLSALDGFNIIIRYQVFTFKNDKGEAIDKGNLVTDS 372
Db 554 -----VPSEENSE-----SVMDGF--VEHPFYATLNGQKYVVMK----- 587

QY 373 KLVLFGRKDDKYEYTGEDKFNVEAIEKDGSMFLDITKPNLSMDKNVFNPSKNNKIYVRNPE 432
Db 588 -----TKDDSYMKDLIIVEGRKVTTVSKDPKN-----NSR 616

QY 433 FYLRGKISDKGGFWELRVNESVVDNLYLYDGLHDNTRFNKLVNKDGMIMDKMKD- 491
Db 617 TLIIPYIPDKAVYNAIVKV---VVANIGYEQYHV---RIINQDINTKDDTSDNTSEP 670

QY 492 -YKANGFPDKVTDMD--GNVYLOQYSDLNAAKAVGVHVFYLYDNVKNPEVNDIPKGNISIE 548
Db 671 LNVQTQGEKGVADTDVAENSSSTATNPKDASDKA-----DVIEPSDVVKDADNNID 721

QY 549 YADGKSVFVNIND-KRANGPDGEIQEHIYINGKEYTSFNDIKQIIDKTLNKKIIVVKDPA 607
Db 722 -KDVQHDVHLSDSMDNNHFD-----KY-----DLKEMDTQ-----IAKOTD 757

QY 608 RNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQF 667
Db 758 RN-----VQKDA-----DNSVGSSNVDTDKD----- 779

QY 668 DGWEISGFEGKADAGYVINLSKDTFIKPVFKKIBEKKEEENKPT-FDVSKKKNPQVNH 726
Db 780 -----SNKNKDKVIQ--LNHIADKNNHTGKAALKDVVKQNYN---NTD 817

QY 727 QLNESHRKEDLQREHS--QKSDSTKQVATVLDKNNIS 763
Db 818 KVTDKKTTEHLPSDIHKTVTKTKKAGTPSKENKLS 856

RESULT 15
US-10-471-571A-1304
; Sequence 1304, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1304
; LENGTH: 1261
; TYPE: PRP

```

```

; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1261)
; OTHER INFORMATION: phi pVL ORF 20 and 21
US-10-471-571A-1304

Query Match      3.7%; Score 147; DB 6; Length 1261;
Best Local Similarity 19.5%; Pred.No.2.9;
Matches 160; Conservative 108; Mismatches 304; Indels 248; Gaps 38;

Qy      1 KLGIEAESK--FKNLGN-----GKEGLAKDITGVVEHHQENEESIKEKSSFTIDRNIS 52
Db      432 KLGITREKALFSELNIFNLISIQHASLSEATELLNSEYLVNDNDLKADLQASLDRAVI- 490
Qy      53 TIRDFENKDLKKLIKFFREVDFTSETGKRME-----EYDYKYDDKGNIIAYDDGDTDL 106
Db      491 -----DVYNOIKNNLESMTPETATIGRLIVDTQALFLEYRKKLQD-----VYTD---- 533
Qy      107 EYETEKLDEIKSKIYGVLSPSKQGHFEILKISNVSKNAKYVYGNMYK-SIEIKATKYDF 165
Db      534 -----VEDVKIAT-----SDRFKLL-----OSQYDEKYKEALEIIATKF-- 568
Qy      166 HSKTMTFDLYANINDIVDGLAFAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYPYVSS 225
Db      569 -----GLTVNEDQLVGEPNVVKSAIAAR--ESTKEQLRDYVKTS 608
Qy      226 Y-----GNV--IELGEGDLKKNKPDNLTKMESKIIYDSEKQYLLKDNILLRGYALKV 278
Db      609 YKTDKGIVERLDTFAEAERTTLKGEIKDKVTLNEYRNGLESEQKYTDDQ-----LSD 660
Qy      279 TTYNPGKIDMEGVNYSKEDIKIQKANPNLRALSETTIYADSR-NVEDGRSTQSVLMS 337
Db      661 LSNNPEIKASTEQANQEAQALKSYIDAQDNLKE-KESQAYADGKISEEQRAIQDA--- 716
Qy      338 ALDGFNIIRVQVTFKND--KGEAIDKGNLVTSSKLVLPKDKDEYTGEDKFVNEAI 395
Db      717 -----QAKLEAKQNAELKARNAEKANAYTDNKV-----KESTDAQR---RTL 757
Qy      396 KEDGSMLFIDTKPVNLSMDKNYFNPSKSNKIYVRNPEFYLRGKISDKGFNWLVRVNESV 455
Db      758 TRYGSQIIQNGKEIKLATTKEEFNATNRLSNILN-----EIVQNV 798
Qy      456 VDNVLI-YGDLHID---NTRDFNKLNVKOGDIMWGMDKYKANGFPDPKV--TMDGNVY 509
Db      799 TDGTTIRYDDNGVAQALNVGPRGIRLNADKIDINGNREINLLIQNMREDKVDKTDIVNSLN 858
Qy      510 LQTCYSDLNKAVGV-----HYQFLYDNVK--PEVNIDPKGNTSYEYADGKSVWFNIN 560
Db      859 LSREGLDINVRIGIKGNNRNVYQIQNDSIELGGIVQRTWKGRKSTD-----DIFTRLK 913
Qy      561 DK-----RNNGPDGEIQBQHI-----YING-----KEYT-----SFNDIK 590
Db      914 DGHLPFRNNTAGGSLYMSHFGISYIDGEGEDGSSGTIQWWDKTYSDSGMNGITINSYG 973
Qy      591 QIDKTL-NKIVVKDPARNTTVE-----FILNKDTGEVSELKPHRVTVTIQNGKMSST 645
Db      974 GVALTSDNNRVVLESYA-SSNIKSQAQPVLYPNTDKVPGL--NRPAFTLSNA----- 1024
Qy      646 IVSEEDFILPVYKGELEKQYQFGWEISGPEGKKDAGYVINLSK----- 689
Db      1025 -----DNAYSDGYITWFGSDENYDYGAGIRFSKERNKGLVQIVNGRYAT 1068
Qy      690 --DTFIKPVFKKIEEKEENKPTFDVSKKKDNPQVNHQS 727
Db      1069 GGDTTIEAGYCK-----FNNLKRDRDGNRYIHQ 1096
```


GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:32:32 ; Search time 21 Seconds
(without alignments)
3541.693 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGEIAESKFNGLNGKEGS.....ATVLDKNNISSKTTNNPNK 773
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2140	2 F95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase,
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypothe
4	250	6.2	1650	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhopty protein -
7	230	5.7	2401	2 T28676	rhopty protein -
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1640	2 A24594	probable major sur
11	218	5.4	1631	1 SAZQK1	major merozoite su
12	218	5.4	1639	2 S05603	major merozoite su
13	217.5	5.4	1546	2 G90603	lipoprotein (impor
14	215.5	5.4	1252	2 B42771	reticulocyte-bindi
15	214	5.3	2339	2 A45597	DNA-directed RNA p
16	212.5	5.3	1125	2 E90598	membrane nuclease,
17	212.5	5.3	2166	2 G70163	hypothetical prote
18	212.5	5.3	2485	1 H71621	serine/threonine-s
19	212	5.3	1188	2 A71621	protein with 5'-3'
20	211.5	5.3	1104	1 A36866	microbial collagen
21	209	5.2	1558	2 B71603	RESA-H3 antigen PF
22	207.5	5.2	4688	2 E89824	hypothetical prote
23	207	5.1	1141	2 E89824	hypothetical prote
24	207	5.1	1191	2 B97116	chromosome segrega
25	207	5.1	1979	2 C71622	hypothetical prote
26	206.5	5.1	1622	2 AE1717	probable cell surf
27	206	5.1	1526	2 A45605	mature-parasite-in
28	204.5	5.1	1516	2 E71619	RAD2 endonuclease
29	204.5	5.1	2500	2 G71609	hypothetical prote

ALIGNMENTS

RESULT 1

F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: F95074
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KUR>
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1,
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0641

Query Match	100.0%	Score	4026;	DB	2;	Length	2140;
Best Local Similarity	100.0%;	Pred. No.	9.5e-166;	Mismatches	0;	Indels	0;
Matches	773;	Conservative	0;				
Qy	1	KLGEIAESKFNGLNGKEGS	LKQTTGV	VEHHHQENEE	SIKEKSSPTIDRN	ISTIRDFENK	60
Db	1334	KLGEIAESKFNGLNGKEGS	LKQTTGV	VEHHHQENEE	SIKEKSSPTIDRN	ISTIRDFENK	1393
Qy	61	DLKLLIKKKFREVD	FTSETG	KRMEEYDYK	YDDKGNIIAYD	DDGTLEYTEKLDEIKSI	120
Db	1394	DLKLLIKKKFREVD	FTSETG	KRMEEYDYK	YDDKGNIIAYD	DDGTLEYTEKLDEIKSI	1453
Qy	121	YGVLSPSKDGHEILGKIS	NSVSKNAKYV	YNNYKSIIEI	KATKYDFHSTMT	FDLYANIND	180
Db	1454	YGVLSPSKDGHEILGKIS	NSVSKNAKYV	YNNYKSIIEI	KATKYDFHSTMT	FDLYANIND	1513
Qy	181	IVDGLAFAGDMRLFVK	ONDQKAEIKR	MPKIKETKSEY	PVSVSYGNVIEL	EGEDLSKN	240
Db	1514	IVDGLAFAGDMRLFVK	ONDQKAEIKR	MPKIKETKSEY	PVSVSYGNVIEL	EGEDLSKN	1573
Qy	241	KPDNLTKMESKIIYSD	SEKQOYLKDN	IIILRGVALK	VTTYNPGKTM	LMLENGVYVSKDI	300
Db	1574	KPDNLTKMESKIIYSD	SEKQOYLKDN	IIILRGVALK	VTTYNPGKTM	LMLENGVYVSKDI	1633
Qy	301	AKIQKPNLRALSETT	IYADSRNV	EDGRSTQSVL	MSALDGFNIIR	YQVFFKMKDKGEA	360
Db	1634	AKIQKPNLRALSETT	IYADSRNV	EDGRSTQSVL	MSALDGFNIIR	YQVFFKMKDKGEA	1693
Qy	361	IDKGNLVTDSSKLV	LFQKDKYEY	TGDKFNVEAI	KEDGSMFLF	IDTKPVNL	SMDKQYNP 420

Dh 1694 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFDITKPVNLSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFWNWLVRNVSVDNYLYIGDLHIDNTRDFNKLNVK 480
Dh 1754 SKSNKIYVRNPEFYLRGKISDKGFWNWLVRNVSVDNYLYIGDLHIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMGMDKYKANGPDKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDYNKVPENID 540
Dh 1814 DGDIMDMGMDKYKANGPDKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDYNKVPENID 1873
Qy 541 PKGNTSIEYADGKSVFVFNINDKRNNGFDGEIQEHYINGKEYTSFNDIKQIDKTLNLIK 600
Dh 1874 PKGNTSIEYADGKSVFVFNINDKRNNGFDGEIQEHYINGKEYTSFNDIKQIDKTLNLIK 1933
Qy 601 IVVKDFARNTTVKSEFILNKDTEVSELSKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Dh 1934 IVVKDFARNTTVKSEFILNKDTEVSELSKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1993
Qy 661 LEKGYPDGEWISGFEKGKADAGYVINLSKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 720
Dh 1994 LEKGYPDGEWISGFEKGKADAGYVINLSKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 2053
Qy 721 PQVNSQLNESHKREDLQREHESQKSDSTKDVTAIVLDKNNISSKSTTNNPNK 773
Dh 2054 PQVNSQLNESHKREDLQREHESQKSDSTKDVTAIVLDKNNISSKSTTNNPNK 2106

RESULT 2
A:97942
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: A97942
R: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A: Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A: Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A: Reference number: A97872; MUID: 21429245; PMID: 11544234
A: Accession: A97942
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-2144 <KUR>
A: Cross-references: UNIPROT: Q8DQP7; UNIPARC: UPI00000E3490; GB: AE007317; PIDN: AAK99365.1;
A: Gene: prtA
C: Genetics:
C: Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 4.1e-152;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLGIASESKFNLGNKSGSLKDDTTGVEHHQNEESIEKSSFTIDRNISTIRDPENK 60
Dh 1339 KLGIASEDKLNAKSPENT--NNQAKDESKPEKSVSEGEASLEINKILISTIREFNK 1397
Qy 61 DLKLLIKKKPREVDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Dh 1398 DLKLLIKKKPREVDFTSETGRMEEDYKYDDKGNIIAYDDGSALEYETEKLDEIKSKI 1457
Qy 121 YGVLSPSKDGHEILGKISNVKAKYVGNVYNNYKSIIEIKATYDFHSTKTMTFDLYANIND 180
Dh 1458 YGVLSPSKDGHEILGKISNVKAKYVGNVYNNYKSIIEIKATYDFHSTKTMTFDLYANIND 1517
Qy 181 IVDGLAFAGDMRLFVKNDQKAEIKTRMPKIKETKSEYYPVSSYGNVIBLGGDLN 240
Dh 1518 IVDGLAFAGDMRLFVKDDNQIKAEIKTRMPKIKETKSEYYPVSSYGNVIBLGGDLN 1577
Qy 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTVTPGKTMLEGNVYSKEDI 300
Dh 1578 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTVTPGKTMLEGNVYSKEDI 1637

Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 360
Dh 1638 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 1697
Qy 361 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFDITKPVNLSMDKNYFNP 420
Dh 1698 IDKGNLVTDSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFDITKPVNLSMDKNYFNP 1757
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFWNWLVRNVSVDNYLYIGDLHIDNTRDFNKLNVK 480
Dh 1758 SKSNKIYVRNPEFYLRGKISDKGFWNWLVRNVSVDNYLYIGDLHIDNTRDFNKLNVK 1817
Qy 481 DGDIMDMGMDKYKANGPDKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDYNKVPENID 540
Dh 1818 DGDIMDMGMDKYKANGPDKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDYNKVPENID 1877
Qy 541 PKGNTSIEYADGKSVFVFNINDKRNNGFDGEIQEHYINGKEYTSFNDIKQIDKTLNLIK 600
Dh 1878 PKGNTSIEYADGKSVFVFNINDKRNNGFDGEIQEHYINGKEYTSFNDIKQIDKTLNLIK 1937
Qy 601 IVVKDFARNTTVKSEFILNKDTEVSELSKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Dh 1938 IVVKDFARNTTVKSEFILNKDTEVSELSKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1997
Qy 661 LEKGYPDGEWISGFEKGKADAGYVINLSKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 720
Dh 1998 LEKGYPDGEWISGFEKGKADAGYVINLSKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 2057
Qy 721 PQVNSQLNESHKREDLQREHESQKSDSTKDVTAIVLDKNNISSKSTTNNPNK 773
Dh 2058 PQVNSQLNESHKREDLQREHESQKSDSTKDVTAIVLDKNNISSKSTTNNPNK 2110

RESULT 3
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C: Species: Melanoplus sanguinipes entomopoxvirus
C: Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C: Accession: T28317
R: Afonso, C.L.; Tullman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A: Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A: Reference number: Z20484; MUID: 99102612; PMID: 9847359
A: Accession: T28317
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 1-1127 <AFO>
A: Cross-references: UNIPROT: Q9YVT6; UNIPARC: UPI00000F6900; EMBL: AF063866; NID: g4049647;
C: Genetics:
A: Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0013;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
Qy 7 ESFKNLGNKSGESLKKDITGVEHHQNEESIEK--EKSSTIID-RNISTIRDPENKDLK 63
Dh 196 EIEFFQINDVQKINQKQDELNKLDESKEFKKQBELNKTIDKQEEELIKLNDKEIN 255
Qy 64 KLKKKPREVDFTSETGRME-----EVDYKYDDKGNIIAYDDGT--DLEYETE 111
Dh 256 FNDEKQKLLDQNSKINTLNENIKGVNLYTETKKNISNLQNEILNKDSTIKSLDEKQK 315
Qy 112 KLDEIKSKIYGVLSPSKDGHEILGKISNVSK-----NAKVYGNVYNNYKSIIEIKAT 161
Dh 316 LLELOKNNITSLVYKSN---TKITNQQLLESSLTDFFNNANI---NINELSKIK 367
Qy 162 KYDFHSTKTMTFDLYANINDIVDGLAFAGDMRLFVK--DNDQKKA-ETKIRMPKIKET-- 216
Dh 368 LFQNDIQKLNNDITEQNNKITD--FFNNSTRIPEKJDTYKIKDDIKNNNLQKLEESYK 425
Qy 217 ----KSEY----PYVSSYGNVIELGEGDLSKNKPDNLTKMESGKIYSDSEKQYLLKDNII 269

Db 426 KIDEQTEYYKINKINKEYNIDIELKNNLQKLEBEENKKIDEQTEYYKINKINKEY--NDIIE 483

Qy 270 LRKGVALKVTTNPG-----KIDMLGNGVYKEDIAKIQKAMPNLRALSETTIYAD 321

Db 484 LKNNLQKLEBEENKINDKLTKLNDIESNTELFNKLNISDFDKSREAKLN--TEYEQ 541

Qy 322 SRN--VEDGRSQSVL-----MSALDGNRIYQVFTFKMDKGEAIDKGNLVTDSSK 373

Db 542 LRKDLLENINKTNELMKLSNKLSSLE-----QLYDSKKNIL-DGIDKIYNSLKEKN- 592

Qy 374 LVLFGKDKKEYGDEKFNVEAIKED--GSMFLFIDTKPVNLSMDKNYFN--PSKSNKIY 427

Db 593 ----DKIDYFNIEKFDIYNWENKPIGNLDSIINKIINNQQKEYINSKIDSKNELS 648

Qy 428 VRNPEFY-----LRKISDKGGFNWELRVNESVVDNYLIYGLHIDNTRD 472

Db 649 TWFDDIFNAKNQIASITNNIENISNKID-----LNEFIISNEDSSKEL-LDEIRK 698

Qy 473 FNIKLVNKGDDIMDMGMKDYKANGPDPKVTMDGNV-YLQTYGSDLNKA-----V 522

Db 699 YKQOFD-KIKDAMNTEVKSFE-NLQKOIDSISINELTNAYDIINTKANLDDKLNNY 756

Qy 523 GVHYQFLYDNVXP-----EWNIDPKGNTSIEYAD-GKSVFVFNINDKRNNGPGBEQH 576

Db 757 GSEFNLVNAGSLDITQKNDEKVKQLNLEYLKNKQSIIEINDIVNN-----805

Qy 577 YINGKEYTSFNDIKQIIDLTKLNKIVKDFARNTTVKFELNKDTGEVSE---LKPFRVT 633

Db 806 FI--KELIKFNTE--TNKSLN-ELLTND--DINDKIFLYKELNKISTNNLLKIY-- 855

Qy 634 VTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDPGWEISGFEKKDAGYVNLKDTF- 692

Db 856 NEIDNVNEKLSIVENLQFINSLSTEFNOG-----SITSH-----INFLNTLA 900

Qy 693 -LKPVPFKLEEKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKD 751

Db 901 GINDVLNKLNLKIMADITTRGTNIR-----DEIKNQISSENISKQKFNKNE--KD 950

Qy 752 VTVATV-----LDKNNISKSTTNN 770

Db 951 LKKLISFNDKLNKYNISAGYTEYN 974

RESULT 4

Ti18444

hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: Ti18444

R:Lawson, D.; Bowman, S.; Barrell, B.

A:Reference number: Z18935

A:Submitted to the EMBL Data Library, August 1997

A:Accession: Ti18444

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1650 <LAW>

A:Cross-references: UNIPROT:O77328; UNIPARC:UPI00000748BD; EMBL:Z98547; NID:e1325376; PI

C:Genetics:

A:Map position: 3

A:Introns: 1597/3; 1625/3

A:Note: C0385c

Query Match 6.2%; Score 250; DB 2; Length 1650;

Best Local Similarity 20.8%; Pred. No. 0.003;

Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

Qy 20 SLKDDTTGVHHQHNEISIKESFTIDRNISTIRDPENKDLKLIKXKPRE---VDDF 76

Db 394 SLNENYNERKIYKEINKEYSKNEV-----HFKND-DSSIKKONSSCLDEQ 443

Qy 77 TSETGKR--MEEDYKYDDKGNIIAY--DDGTDLE---YETKLEIKSKIYGVLSPSKD 129

Db 444 KKKTYKTYIEQRYNFNDRDNNNAYIKDDTHKKEGYLLNMIQOSEYKRYG--SNNKM 501

Qy 130 GHFEILGKIS-----NVSQNAKVYGNKYSI--EIKATKYDFHSKTTFDLYANINDIV 182

Db 502 DEMEYLNQHTNDFNINENLNKIYF-DDYEGYDPEKKKKLDDHLYTQOKEYKNNINIL 560

Qy 183 DGLAPAGDMRULFVKNDOKKAEIKIRMPKIK-----ETKGEYPVYSSVGNVIEGEG-- 235

Db 561 K-----DHLLNDKETKEKQNEIEIEEEKKNKIEIEEEKKNKIEIEEEKKNKIEIEEEKK 614

Qy 236 ---DLSKNKPNLTKMESGKIYSOSEKQOYLKKNII-----LR 271

Db 615 KKIEIEEEKKKKI-EMEEENKIIDDEKKNYANDKIITSHIDNVNINIKIDALLDHIIEKK 673

Qy 272 KGYALKVTVYTPNGKTD-----MLEGNGVYSKEDIAK-----IQKANPNL-- 310

Db 674 KTGHKEINLYKEIKNEYQKMLNDENSIMLEHEKKYNTHQVNNNLCDTKMDLOKENKILTN 733

Qy 311 RALSETTIYADSRNVEDGRSTQSVMSALDGNFIIRYQVFTFKMDKGEAIDKGNLVTD 370

Db 734 NDKKKTFLLSKSKNI-----TSNVLSSKIPG-----TLSTKKNLNTIKTKD--VTD 779

Qy 371 SSKLVLPF-----KDDKE-----YTGEDKF-NVEA-----IKEDGSMFLFIDTK 407

Db 780 NEKKKYVHDHKKNIILKRNKEFINIYKGRNYANVEIGSEVCNNKNVYKGGDNKQMVENK 839

Qy 408 P-----VNLSDMKNYF-----NPSKSNKIYVNPPEF 433

Db 840 QGDDNNMVENKQGDENNVIKNIERSESSPMFTHRRKNITSTDTCAKNEQIKYPHIL 899

Qy 434 YLRKISDKGFGNWEIR-VNESVVDNY-----LIYGLHIDN-----469

Db 900 Y-SNKNEKDKKSIPLKINENIKKNYKDEKISTLEKKVFKKDNVITNNDDEKHTSSK 958

Qy 470 -TRDPNI-----KLN-----478

Db 959 INDDFNITIDKQKGLNPNVDLNRKIKNETYKILEKKSHSKIQNNLMKKKTPNPTNNK 1018

Qy 479 -VKGDIDMDGMKDYKANGPDPKVTMDGNVYL-----QTG 513

Db 1019 GISSTSSSSSTKNFKDCGIIIEKNKLNLSKYTCIRKNQNVDSIKLNDKADLYDKDKKTS 1078

Qy 514 YSDLNKAVGVHYQFLYDNVNVKPEVNIIDPKGNTSIEYADGKSVFN-----INDKRNNGFDG 569

Db 1079 FNDINRAAKGMNFK-----KRDV---PNKNMVDTNKGKR-VFNVPVTLNNYRN--- 1124

Qy 570 EIQQHLYINGKEYTSFNDIKQIIDLTKLNKIVKDFARNTTVKFELNKDTGEVSELKP 629

Db 1125 -----YIRSNK-----NNVK-----NGKMV-----GIKKIVLLKE--KQKSLHP 1156

Qy 630 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDPGWEISGFEKKDAGYVNLKSK 689

Db 1157 EGVEADKKLNSYNDKYLIEKDGFKDIINEEMEK-----YKNNK-MKYKI---K 1201

Qy 690 DTFIKVPFKKIEIEEEKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREE---HSQKS 746

Db 1202 SNSIPIIKKIERKSNNDN---DNIKNDNDINSNNNSNNDKCLFLSEKEDRVHLKNN 1258

Qy 747 DSTKDVTATVLDKKNISKSTTNNPNK 773

Db 1259 NIIVNNTMMFRKQSCDNNTTSLKNK 1285

RESULT 5

Ti18427

hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: Ti18427

R:Lawson, D.; Bowman, S.; Barrell, B.

A:Reference number: Z18935

A:Accession: Ti18427

A:Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA	
A;Residues: 1-3724 <LAW>	
A;Cross-references: UNIPROT:O77320; UNIPARC:UPI000007935D; EMBL:Z98547; NID:e1325376; PI	
C;Genetics:	
A;Introns: 307/1; 1545/2	
A;Note: C0335C	
Qy	5 IAEKFPNLGNGKEGS--LKK-----DTTGVEHHHQENBSIK-----EKSSFT 46
Db	249 IPKNLFENLLNNKQNDYLRNIILMDVNDINLPHEQDQNSLNKKNKCLTGTNKKKEKYM 308
Qy	47 I-----DANISTIRDFENKDLKLIKKEFREVDVDFSTGRMEEDYKYDDKG 95
Db	309 IPKGTEYQDKKESSILTNQNDKKYKK--KKYSELDQSDNISSNNTLSKRYTYT--- 363
Qy	96 NIAYDDGTDLEYTEKLEDEIKSIYGVLSPSGDGHEIFLGKISNVSKNAKVYVGNKYK 155
Db	364 -----CGMDKETKETDEQNK-----RNTEVASICNVCNVTNEKQRKTKTKKK 407
Qy	156 IEIKATKYDFHSKTMTFDLYANINDIVDGLAFAGDMR--LFVKNDQKAEIKRMP-EX 212
Db	408 ---SVQKDDLVEENALLD---NL-IIIDGINFDDVTREKLIIDNNENIDNNIYVENNK 460
Qy	213 IKETKSYPPYVSSYG-NVI-----ELGEGDLSKNKPDNLTQMESGK 252
Db	461 LKQDKS-YDLFSSSGKNLILGVNGBEEFNEVFENIEKLEQERKDEKNKDK-----T 512
Qy	253 IYDSEKQOYLLKDNIIILRGYALKVTTYTPGKTDMLGNGVYSKEDIKIQKAN-PNLR 311
Db	513 IYNNNEQTDLNRNI-----NKIESINNNDN---NNNINKKEFNKIRTEHILNKE 562
Qy	312 ALSETTIYADSRNVDGSRSTQSVLMSALDGFNIIRYQVTFKMDKGEAIDKQGNLVTD 371
Db	563 SISXHIKGPSRDKE-----IKLUTYNNKEDSTFELKLELEIITNN 604
Qy	372 SKLVLFQKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVLNLSMDKNVFNPSKSN----- 424
Db	605 KVNYY--EEDIIGSNEDEYHVLKE-----NLKEDANEYNDKENKNKTKE 650
Qy	425 ----KIYVRN-----PFYLRGKTSKGGFNMELRVN---ESVVDNLYIYDGLHIDNTRD 472
Db	651 ILKSKVLENEKRTLEELKLRGK---NNIFKKDEKYNLSGVEIINEIQINEENKINDIOD 707
Qy	473 FNIK-----LNVKD---GDIMDGMKDYKANGFPDKVTDMD----- 505
Db	708 GNISKQKIIOSSRTNDFNFKDISLNDLLEKEKKRKKSQHFIDNLVKADQNEISENINK 767
Qy	506 -----GNVYLQT-----GYSDLNAKAVGVHY----- 526
Db	768 ICDNINNIIYDESINNIIYDESINNIIYDESINNIIYDENINNIIYDENINNIIYDE 827
Qy	527 -----QFLYDNVNVKPEVID-----PKGNTSTIE-----Y 549
Db	828 NINNIIYDEGINIKICDDNILENKIKTTNDIIVQVEENNESEIEKMLSLNKDINNNTYMF 887
Qy	550 ADGKSVFVN-----INDKRNNGFDCIEIOE----- 573
Db	888 KENVDFIFNKIKRSLKIDKNIKDNNNDIYIMDNFVENDFINHKMEITNKELDPLEI 947
Qy	574 -----QHIYINGKEYTS---FNDIKQIIDKTLNIIKIVKDFARNTTTVKEFILNKDTGE 623
Db	948 NTQNEFIENLOJ KKKKYTNDFHDADQMFYE--NN-KILNKDMKKNEQEPEFKTDETFGS 1005
Qy	624 VSELK-----PHRVTVTTQNGKEMSSSTIVSE 649
Db	1006 LQSHKIKKYNKGBEEKHDKNNEEERNILYDENQVSVLYSDHKIEQDIDQIHSIQTNICDE 1065
Qy	650 EDFILPVYKLEKGYQPDGWEISGFEKGKDAGVYNLSKDTFTFKPVFKKEEKEENK 709

Db	1066 NN-IEQINEENSKKGVIRISGTDN---ENKND-----MENKNDMEKK 1102
Qy	710 PTFDVSXKXNDPQVNH-SQINSHRKKD-QREEHSQKSDSTKDVTAIVLDDKNNISKSTT 768
Db	1103 N--DMEKNDIEKNDMEKKNDMEKKNDMEKKNDMEKKNDME-----NENNENKSDI 1153
Qy	769 NNPKN 773
Db	1154 ENENK 1158
RESULT 6	
T28677	
rhoptry protein - Plasmodium yoelii	
C;Species: Plasmodium yoelii	
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004	
R;Accession: T28677; C45521	
R;Keen, J.; Sinha, K.; Brown, K.; Holder, A.	
Mol. Biochem. Parasitol. 65, 171-177, 1994	
A;Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.	
A;Reference number: Z20508; MUID:95021522; PMID:7935623	
A;Accession: T28677	
A;Status: preliminary; translated from GB/EMBL/DBJ	
A;Molecule type: DNA	
A;Residues: 1-2269 <KE>	
A;Cross-references: UNIPROT:Q26223; UNIPARC:UPI0000080871; EMBL:L27838; NID:9457145; PID	
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.	
Mol. Biochem. Parasitol. 42, 241-246, 1990	
A;Title: Identification of the gene for a plasmodium yoelii rhoptry protein. Multiple co	
A;Reference number: A45521; MUID:91101660; PMID:2270106	
A;Accession: C45521	
A;Status: preliminary	
A;Molecule type: DNA	
A;Residues: 2131-2269 <KE2>	
A;Cross-references: UNIPARC:UPI000017B646; GB:M34283	
Query Match 5.8%; Score 231.5; DB 2; Length 2269;	
Best Local Similarity 19.7%; Pred. No. 0.028;	
Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;	
Qy	37 ESIKESSTIDRNIISTIRDFENK-DLKKLIKKEFREVDVDFSTGRMEEY----- 87
Db	109 EGLKLELKKIKDIIAKIEYVKNVTELKKEIKNNAYIDELANQSPYKVTGYIENKNTIY 168
Qy	88 -----DYKDDKGNIIAYDDGTDLEYTEKLEDEIKSIYGVLSPSKDGCHP 132
Db	169 NTIKSYFDQIYEGDIDTFYNELSSIVKEDPIDIEDKT-KLENLRSKIDNVYDKIQMEI 227
Qy	133 E-ILGKISNVSKNAKVYVGNVYKSIIEIKATKYDFHSKTMTFDLYANINDIVDGLAFAGDM 191
Db	228 ETYKSHLNNIETNNKL-----PNTILIKKIYDEISK-----ELNKMLEDFNK---EKEL 276
Qy	192 RLFEVNDQDKK---AEIKRMPE-----KIKETKSEYPYVSSYGNVIELGE 234
Db	277 SNKISDYDKKREQLESEYKSKMLEIRUHYNSQTVNDNTKKEEAKQNVYDKSNEHMTTIPNE 336
Qy	235 GDLSK-----NKPDLTKMES-----GKIYSD---SEKQOYL-LKQNI---ILRKGYAL 276
Db	337 DEISKIISSEYKTMKDEILSKVNTYIDPNKKYKETVNSEHSQFTELTDKIKAEVSDKELKK 396
Qy	277 KVTTPNGKTDMLGNGVYSKE--DIAKIQKANPNLKALSET-----TIYAD-- 321
Db	397 CEQSFNDKSLINETKNSIEKEYQNINTLKKVDEYIKVCKSTKESITKPSKQTLKXML 456
Qy	322 SRNVDEGRSTQSVLMSALDGFNIIRYQVFT-----FKMN-----DKGEAI----- 361
Db	457 NQNIKTVKETNSDKSVIEKFE---QILTGKQTKLENKFTFSLNHNHEANNNELIKYFS 512
Qy	362 DKQGNLVDSKSLVLFQKDKKEYTGEDKFNVEAIKEDGSMLEFIDTKPVLNLSMDKNVFNPS 421
Db	513 DLKANLGINFENMLYNOQFTEKEKTFND-----IKEQNIHINEEISKIRKIHASINIS 566
Qy	422 KSNKIYVRNPEFYLRGKISDKGGFNBELRVNESVVDNLYIYDGLHIDNTRDFNKLNVK- 480

Db 746 IBSRSISQGNHNEIYGASIKCAQGVDRDNKLDGNHNDKKDDNKDGD---DKKDDNKD 802

Qy 57 FENKDLKLLIKKKPREVDDFTSEFGKMEBYDYKIDKGNIIAYDDGDTLEYE----- 109

Db 803 GDKKDDDKGDKDYDDGD-----KYDDDDNKYDDDDI--YDDDDDFNFDHSETSK 853

Qy 110 -----TEKLDEIK-SKIYGVLSPSKDGHEILGKIS-NVSKNAKYVG 150

Db 854 RLSNYDSLLKKKKGSLKKBEDTLKCSKMINVOEKEKSKMYCMTWNCNVKE-KNYFV 912

Qy 151 NNYKSIKATKYDFHKTMTFDLIYANINDIVDGLAPAGD-----MRLF-----VKON 198

Db 913 NDYMHIIINLSDHEMNINMKSPGYNLFXDVEHFIHKRDFINLILKLIKELIIVQLQN 972

Qy 199 DQKAEIKIRMEPIKETKSEYVVSYGNV-----IELGEGDLS-----KN 240

Db 973 NYKQVCVCKTTGK-TKCMFPSCSTYFHIYCYNKYMOHVRKKNDLQCASQKSRKN 1031

Qy 241 KPDNLTKWE--SGKIYSDSEKQVLLKDNILIRGYALVTTYNPCKTDMLEGNGVYSKE 298

Db 1032 QRAASVRRGAEGKDKGGRSLDRKKDKVTRRSNKYEESANDSGKKN--NGDDKKKG 1089

Qy 299 DIAKIQANPLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVFTFKMDKG 358

Db 1090 NIGKNNNNNNDDSDNNNGDDSDNNNG-----DGNNINGDGNNINGDGN 1143

Qy 359 EAIKDGNLVT-----DSSKLVLFQKDKKEYTGDKFPNV 392

Db 1144 NNINGDGNNDNCLLEKKNQGNQKRRKRNRGRNNDDTSSLVLLNKDDSRQKKKNYG- 1202

Qy 393 EAIKEDGSMPLDITKPVNLSMDKNYFNPSKSNKIYVNPFPYLRGKI-SDKGGFNWELRV 451

Db 1203 -----KDHS-----DQKDSFKSNRRSKN-----KVKDKVEGN-----I 1234

Qy 452 NESVVDNLYIGDLHIDTRDFNKLNVKXGD-----IMDWGMKYKANGFPDKVTDMDGN 507

Db 1235 NDDKINN-----NTDNDKNNNI--NGDNKNNNINGDNKNNING--DKI---NNN 1278

Qy 508 VYLOTGYSDLNAAKAVGHY-QFLYDVKVPEVNDIPKNTSIEYADGKSVVFNINDK----- 562

Db 1279 INGDKINNINGDKISKNYHSHYDN--EETHEQKNFN-KNKNTKRYKRSINDKEMND 1335

Qy 563 -----RNGFDGE-----IOEOHIYINGKEVTS-----FNDIKQIIDKTLNIKIVVKDPA 607

Db 1336 ILNNKNDTDKLNKNYFLMEYQKIISSDKITSGISNMKDINKIKO-----IKDIK 1387

Qy 608 RNTTVKEFILNKDTGEVSELKPHRTVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQF 667

Db 1388 DIKNIKI---KDIKIDIKIDIKIKVKNIKD-----IKSNED----- 1424

Qy 668 DGWEISGFEKKDAGYV-----INLSKDT--FKPVFKTIEE--KKEENKPTFDVSKKK 718

Db 1425 ----IKDIKDVNDVSSVHLNKCVMNKSQDQFCPSDDKNINIIKKNKINIKTDQVY--- 1477

Qy 719 DNPVNSQLNESHKEDLQREHSQKSD-STKDVTATVLDKNNISSKSTTNPN 772

Db 1478 -NPLVDSSTSDCNKNYKKEKAVKKGIGNIKIKLHNNNNNNNNNNNNNNNN 1531

RESULT 9

T18429

hypothetical protein C0345w - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T18429

R;Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A;Reference number: Z18935

A;Accession: T18429

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1711 <LAW>

A;Cross-references: UNIPROT:O77322; UNIPARC:UPI000007873B; EMBL:Z98547; NID:e1325376; PJ

C;Genetics:

A;Note: PFC0345w

Query Match 5.5%; Score 221.5; DB 2; Length 1711;

Best Local Similarity 19.7%; Pred. No. 0.053;

Matches 160; Conservative 134; Mismatches 274; Indels 243; Gaps 37;

Qy 29 EHHQNESESIEKSS-----FTIDRNISTIRDFENKDLK-----KLIKKKPREVD 74

Db 984 DNNDDNNNNMYDSSSSNNHYILTNDRKLN-MDNFINNNINNSQNKVIEKNLEYIN 1042

Qy 75 DF-TSSTGRMEBYDYKIDKGNIIAYDDGDTLEYETEKLEIKSKIYGVLSPSKDGHE 133

Db 1043 NVKLTITS-----NYEQSNNTNSKDE-----HNISDSKSKEDTLNLSRKSSE 1086

Qy 134 ILGKI--SNVSKNAKYVYGNKYKSIKATKYDFHKTMTFDLIYANINDIVDGLAPAGM 191

Db 1087 YNNKILQSTNSKSLANGAYENNLPSGKKKNK-----GTVLKDI-EHINDIQD--KYPEDL 1138

Qy 192 RL-----FVKDNDQKAEIKIRMEPIKETKSEYVVS-----YGNVIELGEGDLSKNKPD 243

Db 1139 NINCNVKYYIENBEKHL-----LPLEL-----EYNLVSSDEKFG-----LNKIKND 1179

Qy 244 NLTKMESGKIYSDSEKQVLLKDNILIRGYALVTTYNPCKTDMLEGNGVYSKEDIKI 303

Db 1180 N-----NIIYMKHQHNLHYDDN---QKHILFDTKN-----VSI 1212

Qy 304 QKANPLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVFTFKMNDK--GEAI 361

Db 1213 QR-NNNINSVIKTNHYEVEKNNKQDN-----YDNFTCDKKKKIYYNII 1255

Qy 362 DKDGNLVTSSKLVLPFGKDKKEYTG-----EDKENVEAIKEDGSMPLDITKPVNL-- 411

Db 1256 NSDKDIYHNN--IITYKNEKEGIGHNLNRNDKDTNFELLKLDGVKBEFLDTPKDSYID 1312

Qy 412 --SMDKNYFNPSKSNKIYVNPFPYLRGKISDKGFNWEIWRVNESVVDNLYIGDLHIDN 469

Db 1313 CHNKKENILNMTNKN-----EDHQIIDVADK-IFN---ETNMITMDNNKIYDDKNVHE 1362

Qy 470 TRDFNKLNVKOGDINDWGMKYKANGFPDKVTDMD--GNVYLOTGYSDLNAAKAVGHYQ 527

Db 1363 KKCTHNDVTHHNNIDILSTSIKNEENLFDITYQKNRIGDIYMNK----- 1407

Qy 528 FLYDVKVPEVNDIPKNTSIEYADGKSVVFNINDKKNNG-----FDGEIOEOHIYINGKEY 583

Db 1408 ---INILQEDDDDDNHN-----NHNNNNNKLIILFEYTKNDQMLHNNKNL 1452

Qy 584 TSPNDIKQIIDKTLNIKIVVKDPAFNTTVKEFILNKDTGEVSELKPHRTVTVTIQNGKEMS 643

Db 1453 EGTEESDFIEKKNKIKIKKNKESYHKIDESLSSNEKNKVSLL-----LINNNKDSS 1505

Qy 644 STTVSEDFILPVYKGELEKGYQFDGWEISGFEKKDAGVIVNLSKDTFIKPVFKTIEEK 703

Db 1506 SV-----DNNKNNNNNNNNNNKNN-----NNNN 1527

Qy 704 KEBENKPTDVSKKDNQPNVNSQLNESHKED-----LQREHSQKSDSTKDVTATV 756

Db 1528 NKKNNKNNNDSFKONNLINNNNNNNNNDSFKONNLINNNNNNNNNNNNNNNNNKVIKEI 1587

Qy 757 LD---KNNTISSKST-----TNPN 772

Db 1588 IDDKENKNDIHKRNIYIKOVSVSPLINHPN 1618

RESULT 10

A24594

probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)

C;Species: Plasmodium falciparum

C;Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 31-Dec-2004

C;Accession: A24594

R;Holder, A.A.; Lockyer, M.J.; Odink, K.3.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, I.

Nature 317, 270-273, 1985

A;Title: Primary structure of the precursor to the three major surface antigens of Plasmo

Qy 551 DGKSVFENIN-----DKRNGFDG---EIQQHIYINGKEYTSFNDIKQIIDKTL 597

Db 215 VLS--FNQBAVKDSFEFYFKEIVSKLKESKINVDKKVDEITLTKTFADF-----SFEQ 267
Qy 175 YANINDIVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYVYVSSYGNVIELGE 234
Db 268 Y--IEKIKDLFKWAST--IKDTYO-----TNKIFLNDY----- 298
Qy 235 GOLSNNKPNLTKME-----SGKYS-----DSEKQYLLKDNILRKGYAL 276
Db 299 --IKKNESNLKKWFSNNFENEIYKYLEKLDALVEIQALHQVLEKNEIIKTGEIVD 356
Qy 277 KYTTPNGKTDLENG--VYSKEDAKI-----OKANPNLRALS 314
Db 357 KVAAFIKSRELKSKAGLLFKSQDSAKITQLINHLISRYFPEAPENLELSKAKMGLINEL 416
Qy 315 ETTIYADSRNVEDGRST--QSVLMSALDGFNIIRYQVTFKMDKGEAIDKGNLVTDSK 373
Db 417 KPEIYVEKEGEIAYPTLQDAISNAQDGQKIFL-----NKNLKLDK--SIVVDKN- 464
Qy 374 LVLFGKDDKEYTGED--KPNVBAIKEDGSMFLFIDTKPNLSMDKNYFNPS----- 421
Db 465 ITIFAKSNVTITRKDSKSFMTFIVQKGALTPEIAEPSQSINLNGLGTSFKDESSLVKI 524
Qy 422 -KSNKIYVNPFPYLRGKISDKGFENWELRVNESVVDNLYIYCDLHIDNTRDFN----- 474
Db 525 EKNAKLVAKGTAFINSKSFYSKYG-----SVFEN---YGSVIEGAKIWNVSESG 572
Qy 475 -----IKLVKDGDIWDGMKDYKANGFPDK-VTDMGNNVLOTGYSDLNNAKAVGVH 525
Db 573 GIIRNHVGSLLTFKNGEI-----RNIISTG--DKGIISQGNIAISGSDGNKS----- 620
Qy 526 YQLFDYDNKPEVNIDPKNGTSTIEYADGKSVFNINDKR-----NNGFDGEIQBHYYING- 580
Db 621 FRSSLINLE-KTNINFNSGSIWNNASVKSILFEIDNSKIQISNNALINPFGSSAIFLKN 679
Qy 581 -----KEYTSFNDIKQIDKTLNIIKIV----- 602
Db 680 STMHLASLEKIKKEASEORIEVLVLPQAKLISPKNIISLDNYQKLSSAIFKIFSVKN 739
Qy 603 VKDFAR-----NTTKETFLNKDT---GEVSELKPHRVTVTIQNGKMSSTIVSEE-DF 652
Db 740 INDFKHVPLVWNTKEKFFKLWPDTKLFVNIFYTKLQNHDLILQSGDFESTEKIIXDELDLF 799
Qy 653 -----ILPVYKGBLEKGYQF-DGWEISGFEGKK-----DAGYVINLSK 689
Db 800 YFRPTAAVKKLILTQVLTIPKYHKFWAEFYPNFDQKWYETINELIRLDPFYLFIAY 859
Qy 690 DTFIKPVPFKIIEKEEENKPTFDVSKKDNQVNHSQLNSHRKEDLQR 739
Db 860 PEFV-----ENGMULPKPEY-IHTNVVNPVLEH-----FRNEDVAR 894

RESULT 14
B42771
reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
C:Species: Plasmodium vivax
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Sep-1997
C:Accession: B42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; PMID:92315338; PMID:1617731
A:Accession: B42771
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1252 <GAL>
A:Cross-references: UNIPARC:UPI000017B645; GB:M88098; NID:g160627; PID:g160628
A:Experimental source: strain Belem, merozoites
C:Genetics:
A:Gene: RBP2

Query Match

5.4%; Score 215.5; DB 2; Length 1252;

Best Local Similarity 20.9%; Pred. No. 0.064;

Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

Qy 2 LGEIAESKFK--NLNGKGEKSLKDDTTGVEHHHOENBESIKEKSSFT-IDRNI----- 51
Db 266 LSEIKYKDKCTTEISNSKRG---KDKTEFLFKPNEESNSKNVINEINENINSEQY 322
Qy 52 -----STIRDFENKOLKLIKKK--FREVDFTS 78
Db 323 LKDIEDIAEQASTKVELPHKHETTISNIFKSEILGVETKSOIKAKAEDIMKEIERHNS 382
Qy 79 ETGRMEBYDYKDDKGNIIAYDDGTLEYTEKLEDEIKSKIYGVLSKDGHFELGKI 138
Db 383 ETQTOVKGFQBNLKNLPHNYDNAED--ELANDKSTNAKVLITETNLESVKHN-----LSEI 437
Qy 139 SNVSNAKAVYCNKYSI--BIKATKYDFHSKTM-----TFDLYAN-INDIVDGLAFAGD 190
Db 438 TNIOGGEKIY-SKAKDIMQIKATSENTAETLEKVKDDQDSYVNYLNOIT--E 490
Qy 191 MRLFVKND-----OKKAEIKIRMPKIKE-----TK-----S 218
Db 491 RNLIIVTEKNRLNGIDSTITNIEGALKESGNYEIGFLEKLEBEIGKRNKLKVDITKKSINS 550
Qy 219 EYPYVSSYGNVIELGEGDLKSNKPNLTKMESKIYSDSEKQOYLLKONIILRKGYALKV 278
Db 551 TVGNFSSLFNFDFLNQYDFNKNINDYENKM--GEIYNEFEGSLNKISENL---RNASENT 605
Qy 279 TTYNPGKTDMLEGNGVYSKEDI---AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVL 335
Db 606 SDYNSAKTLRLEA--QKEKVNLLNKEEANKYLR-----DVKKVESFR----- 646
Qy 336 MSALDGFNIIRYQVTFKQNDKGEAIDKGNLVTDSKLVFGKDDKEYTGBDKFNVEAI 395
Db 647 -----FIFNMK--ESLDKINEMIKKEQLTV-----NEGHGNVKQLVNI 683
Qy 396 KEDGSMFLIDTKPNLS-----MDKN-----YFNPSKSNKIYVRNPEFYLRGKI 439
Db 684 KE-----LVDEN--NLSDLKQATGKNBEIOKITHSTLNKAKTILGHVDTSAKYVGKI 736
Qy 440 SDKGNNFELRVNESVVD-----NLYIYDGLHIDNTRFNKLNKVD--- 481
Db 737 TP-----ELALTGDLGAKLTAQELKFESKNNVLETNMKNKTNELDVHKNIQDAYK 790
Qy 482 -----GDIMDMGMKDYKANGFPDKVTDMDGNVYL-----QTYGVDLNAKAVGVH 525
Db 791 VALEILAHSDSIDTKQD-----SSKLIEMGNQIYLVKVLINQYKNKISSIKSEEAVS 844
Qy 526 YQFLYDNVK-----PEVNIDPKNGTSTIEYADGKSVFNI-----NDKRNNGFDGEIQEQ 574
Db 845 VKI--GNVSKKHSBELSKITCSKSYDNIIALEKQTELQNLNRSFTQKNTNTNSDKLE-- 900
Qy 575 HIYINGKEYTSFNDIKQIDKTLNIIK-----VVKD--- 605
Db 901 -----KIKTDFESIKNAL-KTLEGEVNALKASDNHEHVQSKSEPVNPALSEIEKETD 953
Qy 606 -FARNTTVKEFILNKDTGEVSELKPHRVTVT--IQNGKMSSTIVSEEDFILPVYKGBLE 662
Db 954 IDSLNTALDELLKGRTECVSRYKLIKQDVTKEISDDDELINTIEKN---VKAYLAVIK 1009
Qy 663 KGYQFDGWEISGFGKKGAGVIVNLSKDTFFKPV-----FKLIEEKEEENKPTFD-- 713
Db 1010 KNYE-----DTVQD---VLTINHFNTKQVSNHPTNFDKSNKSEBELTKAVTDSK 1057
Qy 714 -VSKKK-----DNQOVN-----HSQLNESHR-KEDLOREHSQ 744
Db 1058 TIISKLKGVILEVNEENTEMTIESAKEIEALYNELNKKTSLSLIEIYOTSNEVKLOEMKS 1117
Qy 745 KSDSTKDVTA---TVLD---KNNISSKSTTN 770
Db 1118 NADKYVIDSVKIFNTVLTQKSNIVTNOHSINN 1149

RESULT 15
A45597
DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodium

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:21:46 ; Search time 96 Seconds
(without alignments)
3681.545 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEIAESKFNGLNGKEGS.....ATVLDKNNISSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_8.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*
- 10: Geneseqp2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	4 AAB48343	Aab48343 S. pneumo
2	4026	100.0	2140	6 ABU01020	Abu01020 S. pneumo
3	4026	100.0	2140	6 ABU45746	Abu45746 Protein e
4	4026	100.0	2140	8 ADM92113	Adm92113 S. pneumon
5	4026	100.0	2140	8 ADT50099	Adt50099 S. pneumon
6	4015.5	99.7	2120	3 AAY81710	Aay81710 Streptoco
7	3789.5	94.1	2138	8 ADK48759	Adk48759 Streptoco
8	3048	75.7	637	8 ADR94534	Adr94534 Novel S.
9	3048	75.7	637	9 AEA58404	Aea58404 Streptoco
10	750.5	18.6	1529	8 ADR96136	Adr96136 Novel S.
11	750.5	18.6	1529	9 AEA60006	Aea60006 Streptoco
12	615	15.3	117	2 AAW55096	Aaw55096 Streptoco
13	615	15.3	117	5 ABP54590	Abp54590 S. pneumo
14	615	15.3	117	7 ADC45149	Adc45149 S. pneumo
15	227	5.6	1196	6 ABU24813	Abu24813 Protein e
16	222	5.5	861	7 ABO23608	AbO23608 Plasmodiu
17	220	5.5	1166	8 ADK99186	Adk99186 Streptoco
18	218	5.4	1639	2 AAW54145	Aaw54145 P. falcip
19	218	5.4	1639	5 AAE29345	Aae29345 Plasmodiu
20	217	5.4	1166	8 ADK99185	Adk99185 Streptoco
21	217	5.4	1166	8 ADK99178	Adk99178 Streptoco
22	217	5.4	1166	8 ADK99184	Adk99184 Streptoco
23	217	5.4	1233	8 ADV82938	Adv82938 Streptoco

ALIGNMENTS

RESULT 1

AAB48343

ID AAB48343 standard; protein; 773 AA.

XX

AC AAB48343;

XX

DT 20-APR-2001 (first entry)

XX

DE S. pneumoniae Spl130 polypeptide.

XX

KW Immunogenic; Spl128; Spl130; pneumococcal; otitis media; nasopharyngeal; bronchial; lung; blood; infection; immune response; immunotherapy;

KW antibacterial; auditory; vaccine.

XX

OS Streptococcus pneumoniae.

XX

PN WO200076540-A2.

XX

PD 21-DEC-2000.

XX

PF 09-JUN-2000; 2000WO-US015925.

XX

PR 10-JUN-1999; 99US-0138453P.

XX

PA (MEDI-) MED IMMUNE INC.

XX

PI Adamou JE, Choi GH;

XX

DR WPI; 2001-112197/12.

XX

DR N-PSDB; AAC84742.

PT

PT New vaccines comprising Spl128 or Spl130 polypeptides, for treating and preventing pneumococcal infections, particularly infections caused by Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or blood infections.

XX

PS Claim 8; Page 51-54; 54pp; English.

XX

CC The invention relates to novel immunogenic polypeptides, Spl128 and Spl130 from S. pneumoniae. Vaccines comprising the polypeptides are useful for the treatment and prevention of pneumococcal infections, particularly infections caused by Streptococcus, such as otitis media, nasopharyngeal, bronchial, lung or blood infections. The antigens are used as immunogenic agents to stimulate an immune response. The antisera and antibodies may also be used in diagnosing and treating pneumococcal infections. Recombinant polypeptides serve as a mechanism for stimulating production

CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the S. pneumoniae Sp130 polypeptide
XX
SQ Sequence 773 AA;

Query Match 100.0%; Score 4026; DB 4; Length 773;
Best Local Similarity 100.0%; Pred. No. 9e-225;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNKLGNGKEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60
Dy 1 KLGEIAESKFNKLGNGKEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60

Qy 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Dy 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120

Qy 121 YGVLSPSKDGHFELIGKISNVSKNAKYVYNNYKSIIEIKATYDFHSHKTMFTFDLIYANIND 180
Dy 121 YGVLSPSKDGHFELIGKISNVSKNAKYVYNNYKSIIEIKATYDFHSHKTMFTFDLIYANIND 180

Qy 181 IVDGGLAFAGDMRLFKVNDQKAEIKIRMPKEIKETKSEYPVSSYGNVIELGEGDLSKN 240
Dy 181 IVDGGLAFAGDMRLFKVNDQKAEIKIRMPKEIKETKSEYPVSSYGNVIELGEGDLSKN 240

Qy 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKGVALKVTTYNPGKTDMLEGNGVYSKEDI 300
Dy 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKGVALKVTTYNPGKTDMLEGNGVYSKEDI 300

Qy 301 AKIOKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNRIQVFTFKMNDKGEA 360
Dy 301 AKIOKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNRIQVFTFKMNDKGEA 360

Qy 361 IDKGNLVTDSKLVLFQKDDKEVTGEDKFNVEAIKEDGSMFLFDTKFPVNLMSMDKNVFN 420
Dy 361 IDKGNLVTDSKLVLFQKDDKEVTGEDKFNVEAIKEDGSMFLFDTKFPVNLMSMDKNVFN 420

Qy 421 SKSNKIYVRNPEFYLRGKISDKGFNMLRVNESVDNLYIGDLHIDNTRDFNKLNVK 480
Dy 421 SKSNKIYVRNPEFYLRGKISDKGFNMLRVNESVDNLYIGDLHIDNTRDFNKLNVK 480

Qy 481 DGDIMDMGMDKYKANGFPDPKVTMDGNYLQGTGSDLNKAVGVHYQFLYDNVKEPVNID 540
Dy 481 DGDIMDMGMDKYKANGFPDPKVTMDGNYLQGTGSDLNKAVGVHYQFLYDNVKEPVNID 540

Qy 541 PKGNTSIEYADGKSVFVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Dy 541 PKGNTSIEYADGKSVFVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600

Qy 601 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQQKEMSSTIVSEEDFILPVYKGE 660
Dy 601 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQQKEMSSTIVSEEDFILPVYKGE 660

Qy 661 LEKGYPFGWISGPEGKDGAGYVNLKSDTFIKPVPKKIEEKEENKPFDFVSKKDN 720
Dy 661 LEKGYPFGWISGPEGKDGAGYVNLKSDTFIKPVPKKIEEKEENKPFDFVSKKDN 720

Qy 721 POVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTTNPNK 773
Dy 721 POVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTTNPNK 773

RESULT 2
ID ABU01020 standard; protein; 2140 AA.
XX ABU01020;
AC ABU01020;
XX
XX 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain protein from coding region #590.

Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 3.5e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNKLGNGKEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60
Dy 1334 KLGEIAESKFNKLGNGKEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 1393

Qy 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Dy 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120

XX
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
XX WO200277021-A2.
PN 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-IB002163.
PF
XX 27-MAR-2001; 2001GB-00007658.
PR
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
XX WPI; 2003-040579/03.
DR N-PSDB; ABX06302.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
PS Claim 1; SEQ ID NO 1180; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise QS field)
XX
SQ Sequence 2140 AA;

Db	1394	DLKKLKKKPREVDDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLDLEYETEKLEDEIKSKI	1453
QY	121	YGVLSPSKDGHFEILGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSHKMTWTFDLYANIND	180
Db	1454	YGVLSPSKDGHFEILGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSHKMTWTFDLYANIND	1513
QY	181	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIEGEGDLSKN	240
Db	1514	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIEGEGDLSKN	1573
QY	241	KPDNLTKMESGKIYSDEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI	300
Db	1574	KPDNLTKMESGKIYSDEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI	1633
QY	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA	360
Db	1634	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA	1693
QY	361	IDKGNLVTDSKLVLPFGKDDKEYTGEDKFNVEAIKEDGSMULFIDTPKPNLSMDKNYFNP	420
Db	1694	IDKGNLVTDSKLVLPFGKDDKEYTGEDKFNVEAIKEDGSMULFIDTPKPNLSMDKNYFNP	1753
QY	421	SKSNKIYVRNPEFYLRGKISDGGFNWELRVNESVVDNLYIYDGLHIDNTRDFNKLNVK	480
Db	1754	SKSNKIYVRNPEFYLRGKISDGGFNWELRVNESVVDNLYIYDGLHIDNTRDFNKLNVK	1813
QY	481	DGDIIMDGMDKYKANGFPDKVTDMDGNVYLQTYSDLNAKAVGVHYOFLYDNVKNPEVNI	540
Db	1814	DGDIIMDGMDKYKANGFPDKVTDMDGNVYLQTYSDLNAKAVGVHYOFLYDNVKNPEVNI	1873
QY	541	PKGNTSIEYADGKSVFVFNIDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNFK	600
Db	1874	PKGNTSIEYADGKSVFVFNIDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNFK	1933
QY	601	IWKDFARNTTVKFEFLNKDTEVSELKPHRVTTVTIQNGKEMSTIVSEEDFILPVYKGE	660
Db	1934	IWKDFARNTTVKFEFLNKDTEVSELKPHRVTTVTIQNGKEMSTIVSEEDFILPVYKGE	1993
QY	661	LEKGQFDGWEISGFEKKDAGVYVNLSDTKFIKPVFKKIEEKKEENKPTFDVSKKDN	720
Db	1994	LEKGQFDGWEISGFEKKDAGVYVNLSDTKFIKPVFKKIEEKKEENKPTFDVSKKDN	2053
QY	721	PQVNHSQLNESHKREDLQREHSQKSDSKDVTATVLDKQNTSSKSTTNPNK	773
Db	2054	PQVNHSQLNESHKREDLQREHSQKSDSKDVTATVLDKQNTSSKSTTNPNK	2106
RESULT 3			
ID	ABU45746	standard; protein; 2140 AA.	
XX	XX		
AC	ABU45746;		
XX	XX		
DT	19-JUN-2003	(first entry)	
DE	Protein encoded by Prokaryotic essential gene #31273.		
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.		
XX	Streptococcus pneumoniae.		
OS	WO200277183-A2.		
PN			
XX	03-OCT-2002.		
XX	XX		
PF	21-MAR-2002; 2002WO-US009107.		
XX	XX		
PR	21-MAR-2001; 2001US-00815242.		
PR	06-SEP-2001; 2001US-00948993.		
PR	25-OCT-2001; 2001US-0342923P.		
PR	08-FEB-2002; 2002US-00072851.		
PR	06-MAR-2002; 2002US-0362699P.		
XX	XX		

PA	(ELIT-) ELITRA PHARM INC.		
XX	XX		
PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,		
PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;		
XX	XX		
DR	WPI; 2003-029926/02.		
DR	N-PSDB; ACA49616.		
XX	XX		
PT	New antisense nucleic acids, useful for identifying proteins or screening		
PT	for homologous nucleic acids required for cellular proliferation to		
PT	isolate candidate molecules for rational drug discovery programs.		
XX	XX		
PS	Claim 25; SEQ ID NO 73670; 1766pp; English.		
XX	XX		
CC	The invention relates to an isolated nucleic acid comprising any one of		
CC	the 6213 antisense sequences given in the specification where expression		
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:		
CC	(1) a vector comprising a promoter operably linked to the nucleic acid		
CC	encoding a polypeptide whose expression is inhibited by the antisense		
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated		
CC	polypeptide or its fragment whose expression is inhibited by the		
CC	antisense nucleic acid; (4) an antibody capable of specifically binding		
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular		
CC	proliferation or the activity of a gene in an operon required for		
CC	proliferation; (7) identifying a compound that influences the activity of		
CC	the gene product or that has an activity against a biological pathway		
CC	required for proliferation, or that inhibits cellular proliferation; (8)		
CC	identifying a gene required for cellular proliferation or the biological		
CC	pathway in which a proliferation-required gene or its gene product lies		
CC	or a gene on which the test compound that inhibits proliferation of an		
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a		
CC	compound's activity; (11) a culture comprising strains in which the gene		
CC	product is overexpressed or underexpressed; (12) determining the extent		
CC	to which each of the strains is present in a culture or collection of		
CC	strains; or (13) identifying the target of a compound that inhibits the		
CC	proliferation of an organism. The antisense nucleic acids are useful for		
CC	identifying proteins or screening for homologous nucleic acids required		
CC	for cellular proliferation to isolate candidate molecules for rational		
CC	drug discovery programs, or for screening homologous nucleic acids		
CC	required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> ,		
CC	<i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of		
CC	the target prokaryotic essential genes. Note: The sequence data for this		
CC	patent did not form part of the printed specification, but was obtained		
CC	in electronic format directly from WIPO at		
CC	ftp.wipo.int/pub/published_pct_sequences		
XX	XX		
SQ	Sequence 2140 AA;		
Query Match			
Best Local Similarity 100.0%; Score 4026; DB 6; Length 2140;			
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	KLGEIAESKFNKLGNGKESLKKDVTGVEHHHQNEESIKEKSFTIDRNIETIRDFENK	60
Db	1334	KLGEIAESKFNKLGNGKESLKKDVTGVEHHHQNEESIKEKSFTIDRNIETIRDFENK	1393
QY	61	DLKKLKKKFFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLDLEYETEKLEDEIKSKI	120
Db	1394	DLKKLKKKFFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLDLEYETEKLEDEIKSKI	1453
QY	121	YGVLSPSKDGHFILGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSHKMTWTFDLYANIND	180
Db	1454	YGVLSPSKDGHFILGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSHKMTWTFDLYANIND	1513
QY	181	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIEGEGDLSKN	240
Db	1514	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIEGEGDLSKN	1573
QY	241	KPDNLTKMESGKIYSDEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI	300
Db	1574	KPDNLTKMESGKIYSDEKQOYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI	1633
QY	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFPMNDKGEA	360

Db 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKWNDKGEA 1693
Qy 361 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1694 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLIYGLDHIIDNTRDFNKLNVK 480
Db 1754 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLIYGLDHIIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMGMDKYKANGPDPKVTDMGNNVLOTGYSDLNKAVGVHYOFLYDNVKEPVNID 540
Db 1814 DGDIMDMGMDKYKANGPDPKVTDMGNNVLOTGYSDLNKAVGVHYOFLYDNVKEPVNID 1873
Qy 541 PKGNTSIEYADGKSVWFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1874 PKGNTSIEYADGKSVWFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 1933
Qy 601 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGE 660
Db 1934 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGE 1993
Qy 661 LEKGQFDPGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Db 1994 LEKGQFDPGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 2053
Qy 721 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNPNK 773
Db 2054 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNPNK 2106

RESULT 4
ID ADM92113 standard; protein; 2140 AA.
XX ADM92113;
AC ADM92113;
XX 03-JUN-2004 (first entry)
DT S pneumoniae antigenic protein sequence SegID310.
XX Streptococcus pneumoniae.
XX WO2004020609-A2.
XX 11-MAR-2004.
XX 02-SEP-2003; 2003WO-US027401.
XX 30-AUG-2002; 2002US-0407082P.
XX (TUFT) UNIV TUFTS.
XX Camilli A, Hava DL;
XX WPI; 2004-239189/22.
XX N-PSDB; ADM91876.
XX New Streptococcus pneumoniae nucleic acid molecules, useful for
PT diagnosing, treating and preventing active infections of Streptococcus
PT pneumoniae.
XX Claim 27; SEQ ID NO 310; 123pp; English.

XX This invention relates to novel isolated Streptococcus pneumoniae nucleic
CC acid molecules and the antigenic polypeptides encoded by them. The
CC invention may be useful for the production of compounds with an
CC antibacterial activity or for gene therapy. The nucleic acid molecules,
CC compositions and methods disclosed are useful for treating Streptococcus

CC pneumoniae infection. The present sequence is that of an S pneumoniae
CC protein of the invention.
XX SQ Sequence 2140 AA;
Query Match 100.0%; Score 4026; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. No. 3.5e-224; Indels 0; Gaps 0;
Matches 773; Conservative 0; Mismatches 0;
Qy 1 KLGEIAESKFNKLGNGKEGSLKDDTTGVEHHHQENEEISIKEKSSFTIDRNIISTIRDPENK 60
Db 1334 KLGEIAESKFNKLGNGKEGSLKDDTTGVEHHHQENEEISIKEKSSFTIDRNIISTIRDPENK 1393
Qy 61 DLKKLKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEYETEKLEDEISKI 120
Db 1394 DLKKLKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEYETEKLEDEISKI 1453
Qy 121 YGVLSPSKDGHFEILGKISNVSKNAKYVYGNKYKSIKATKYDFHSHKTWTFDLYANIND 180
Db 1454 YGVLSPSKDGHFEILGKISNVSKNAKYVYGNKYKSIKATKYDFHSHKTWTFDLYANIND 1513
Qy 181 IVDLAFAGDMRFLVKDNDQKAEIKIRMPKIKETKSEYYPYVSSYGNVIELSGEDLSKN 240
Db 1514 IVDLAFAGDMRFLVKDNDQKAEIKIRMPKIKETKSEYYPYVSSYGNVIELSGEDLSKN 1573
Qy 241 KPNDLTKMESGKIYSDSEKQOYLLKUNIIIRKGYALKVTTYNPGKTDMLSGNGVYSKEDI 300
Db 1574 KPNDLTKMESGKIYSDSEKQOYLLKUNIIIRKGYALKVTTYNPGKTDMLSGNGVYSKEDI 1633
Qy 301 AKTQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKWNDKGEA 360
Db 1634 AKTQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKWNDKGEA 1693
Qy 361 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1694 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLIYGLDHIIDNTRDFNKLNVK 480
Db 1754 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLIYGLDHIIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMGMDKYKANGPDPKVTDMGNNVLOTGYSDLNKAVGVHYOFLYDNVKEPVNID 540
Db 1814 DGDIMDMGMDKYKANGPDPKVTDMGNNVLOTGYSDLNKAVGVHYOFLYDNVKEPVNID 1873
Qy 541 PKGNTSIEYADGKSVWFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1874 PKGNTSIEYADGKSVWFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 1933
Qy 601 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGE 660
Db 1934 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGE 1993
Qy 661 LEKGQFDPGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Db 1994 LEKGQFDPGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 2053
Qy 721 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNPNK 773
Db 2054 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNPNK 2106

RESULT 5
ID ADT50099 standard; protein; 2140 AA.
XX ADT50099;
XX AC ADT50099;
XX

DT 13-JAN-2005 (first entry)
XX S_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.
XX hyperimmune serum reactive antigen; antibacterial; vaccine;
KW

KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;
KW sepsis; meningitis.
OS Streptococcus pneumoniae TIGR4.
XX WO20040922209-A2.
PN 28-OCT-2004.
PD 15-APR-2004; 2004WO-EP003984.
XX 15-APR-2003; 2003EP-00450087.
PR (INTE-) INTERCELL AG.
XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;
PI WPI: 2004-758335/74.
DR N-PSDB; ADT49955.
XX New hyperimmune serum reactive antigens from Streptococcus pneumoniae,
PT and encoding nucleic acid molecules, useful for diagnosing, preventing or
PT treating S. pneumoniae infections.
XX Disclosure; SEQ ID NO 177; 191pp; English.
PS This invention relates to novel nucleic acids encoding hyperimmune serum
XX reactive antigens, or fragments derived thereof. Specifically, it refers
CC to antigens selected from peptides and serum reactive epitopes that can
CC be used in pharmaceutical compositions that exhibit antibacterial
CC activity. The present invention describes a composition (including the
CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)
CC that is useful for manufacturing a medicament such as a vaccine, which
CC can be used to treat or prevent bacterial infections, particularly S.
CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,
CC bacteraemia sepsis and meningitis. The antigen or its fragment may also
CC be used for isolating, purifying and/ or identifying an interaction
CC partner of the hyperimmune serum reactive antigen, as well as for
CC manufacturing a functional nucleic acid selected from aptamers and
CC spiegelmers or for manufacturing a functional ribonucleic acid selected
CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide
CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen
CC of the invention.
XX Sequence 2140 AA;
SQ Query Match 100.0%; Score 4026; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. NO. 3.5e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLGEAIESKFKNLGNGKESLKKDTTGVHHQHEEESIKESSTFIDRNISTIRDFENK 60
DB 1334 KLGEAIESKFKNLGNGKESLKKDTTGVHHQHEEESIKESSTFIDRNISTIRDFENK 1393
QY 61 DLKLLIKKFKREVDDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEYETEKLEIKSKI 120
DB 1394 DLKLLIKKFKREVDDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEYETEKLEIKSKI 1453
QY 121 YGVLSPSKDGHFEILGKISNVSKNAKYVGNKYKIEIKATKYDFHSTKMTFDLYANIND 180
DB 1454 YGVLSPSKDGHFEILGKISNVSKNAKYVGNKYKIEIKATKYDFHSTKMTFDLYANIND 1513
QY 181 IVDGLAFAGDMFLFKVNDQKAEIKIRMPKIKETKSEYPVSVYGNVIELGEGDLKN 240
DB 1514 IVDGLAFAGDMFLFKVNDQKAEIKIRMPKIKETKSEYPVSVYGNVIELGEGDLKN 1573
QY 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
DB 1574 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633
QY 301 AKIQANPNLRALSETTYIADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTPKMNDKGEA 360
DB 1634 AKIQANPNLRALSETTYIADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTPKMNDKGEA 1693

QY 361 IDKGNLVTDSKLVLFKGDKEYTGEDKFNVAEIKEDGSMFLFIDTKPVLNSMDKNYFNP 420
DB 1694 IDKGNLVTDSKLVLFKGDKEYTGEDKFNVAEIKEDGSMFLFIDTKPVLNSMDKNYFNP 1753
QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLIYDGLHIDNTRDFNKLNVK 480
DB 1754 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLIYDGLHIDNTRDFNKLNVK 1813
QY 481 DGDIMDWGMKDYKANGFPDKVYTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEVNI 540
DB 1814 DGDIMDWGMKDYKANGFPDKVYTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEVNI 1873
QY 541 PKGNTSIEYADGKSVWFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDTLNKK 600
DB 1874 PKGNTSIEYADGKSVWFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDTLNKK 1933
QY 601 IVVKDFARNTTVKPEILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
DB 1934 IVVKDFARNTTVKPEILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1993
QY 661 LEKGYQFDGWEISGPEGKKDAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSKKDN 720
DB 1994 LEKGYQFDGWEISGPEGKKDAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSKKDN 2053
QY 721 PQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
DB 2054 PQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106
ID AAY81710 standard; protein; 2120 AA.
XX AC AAY81710;
XX DT 02-JUN-2000 (first entry)
XX DE Streptococcus pneumoniae protein sequence ID3.
XX KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
XX OS Streptococcus pneumoniae.
XX PN WO200006738-A2.
XX PD 10-FEB-2000.
XX PF 27-JUL-1999; 99WO-GB002452.
XX PR 27-JUL-1998; 98GB-00016336.
XX PR 19-MAR-1999; 99US-0125329P.
XX PA (MICR-) MICROBIAL TECHNIQS LTD.
XX PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX DR WPI: 2000-195301/17.
XX DR N-PSDB; AAZ91806.
XX PT Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections.
XX PS Claim 2; Page 41-42; 76pp; English.
XX CC This sequence represents a Streptococcus pneumoniae protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or

CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC with chem. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of S. pneumoniae
CC infection. As the sequences can be used to treat S. pneumoniae infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis
XX
SQ Sequence 2120 AA;

Query Match 99.7%; Score 4015.5; DB 3; Length 2120;
Best Local Similarity 99.9%; Pred. No. 1.4e-223;
Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 KLGEIAESKFNKLGNGEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60
Db 1313 KLGEIAESKFNKLGNGEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 1372
Qy 61 DLKKLIIKKKPREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVEYETKLDIEIKSKI 120
Db 1373 DLKKLIIKKKPREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVEYETKLDIEIKSKI 1432
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIEIKATKYDFHSTMTFIDLYANIND 180
Db 1433 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIEIKATKYDFHSTMTFIDLYANIND 1492
Qy 181 IVDGLAFAGDMRLFKVNDQKABIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN 240
Db 1493 IVDGLAFAGDMRLFKVNDQKABIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN 1552
Qy 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1553 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVTTYNPGKTDMLGNGVYSKEDI 1612
Qy 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA 360
Db 1613 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA 1672
Qy 361 IDKGNLVTDSKLVLFKGDDEYTGEDKFNVEAIKEDGSMFLFDTDKPVLMSMDKNYFNP 420
Db 1673 IDKGNLVTDSKLVLFKGDDEYTGEDKFNVEAIKEDGSMFLFDTDKPVLMSMDKNYFNP 1732
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFFNWLVRNVSVDNLYIYDGLHIDNTR - DFNIKLVN 479
Db 1733 SKSNKIYVRNPEFYLRGKISDKGFFNWLVRNVSVDNLYIYDGLHIDNTRDDFNIKLVN 1792
Qy 480 KGDIMDMGMDKYKANGPPDKVTMDGNVYLQTYGSDLNAKAVGVHYQFLYDNNVKEPVI 539
Db 1793 KGDIMDMGMDKYKANGPPDKVTMDGNVYLQTYGSDLNAKAVGVHYQFLYDNNVKEPVI 1852
Qy 540 DPKGNSTYADGSKVFNINDKRNNGPDGIEIQHIYNGKEYTSNDIKQIIDKTLNI 599
Db 1853 DPKGNSTYADGSKVFNINDKRNNGPDGIEIQHIYNGKEYTSNDIKQIIDKTLNI 1912
Qy 600 KIVVKDFARNVTVEFLNKDTGVESELKPHRVTTVIONGKEMSTIVSEEDFILPVYKG 659
Db 1913 KIVVKDFARNVTVEFLNKDTGVESELKPHRVTTVIONGKEMSTIVSEEDFILPVYKG 1972
Qy 660 ELEKGYQDGEWIEISGFEKGKAGVYVINSKDTFKIPVFKKIEBKKEENKPTFDVSKKKD 719
Db 1973 ELEKGYQDGEWIEISGFEKGKAGVYVINSKDTFKIPVFKKIEBKKEENKPTFDVSKKKD 2032
Qy 720 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISKSSTTNPNK 773
Db 2033 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISKSSTTNPNK 2086

RESULT 7
ADK48759

ID ADK48759 standard; protein; 2138 AA.
XX
AC ADK48759;
XX
DT 20-MAY-2004 (first entry)
XX
DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX
KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
PN US6699703-B1.
XX
PD 02-MAR-2004.
XX
PF 26-MAY-2000; 2000US-00583110.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CB;
XX
DR WPI; 2004-212399/20.
DR N-PSDB; ADK46098.
XX
PT New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2138 AA;

Query Match 94.1%; Score 3789.5; DB 8; Length 2138;
Best Local Similarity 94.8%; Pred. No. 1.8e-210;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;
Qy 1 KLGEIAESKFNKLGNGEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60
Db 1333 KLGEIPESKFNKLVKVDLSLNKETAENVNLLVDNQSIKESKSLFNHKTISTIRDENK 1392
Qy 61 DLKKLIIKKKPREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVEYETKLDIEIKSKI 120
Db 1393 DLKKLIIKKKPREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVEYETKLDIEIKSKI 1451
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIEIKATKYDFHSTMTFIDLYANIND 180
Db 1452 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIEIKATKYDFHSTMTFIDLYANIND 1511
Qy 181 IVDGLAFAGDMRLFKVNDQKABIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN 240
Db 1512 IVDGLAFAGDMRLFKVNDQKABIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN 1571
Qy 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1572 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVTTYNPGKTDMLGNGVYSKEDI 1631
Qy 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA 360

Db 1632 AKIQKANPNLRALETTIYADSRNVEDGSRSTQAVLMSALDGFNIIRYQVFTFMNDKGEA 1691
QY 361 IDKDGNLVTDSSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 420
Db 1692 IDKDGNLVTDSSKLVLFQKDDKEYTGEDKSNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 1751
QY 421 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNLYLYGDLHDNTRDFNKLNVK 480
Db 1752 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNLYLYGDLHDNTRDFNKLNVK 1811
QY 481 DGDIMWGKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEPVNID 540
Db 1812 DGDIMWGKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEPVNID 1871
QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSNDIKQIIDKTLNLIK 600
Db 1872 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSNDIKQIIDKTLNLIK 1931
QY 601 IVVKOPARNTTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKGE 660
Db 1932 IVVKOPARNTTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKGE 1991
QY 661 LEKGYQFDGWEISGPEGKXDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDN 720
Db 1992 LEKGYQFDGWEISGPEGKXDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDN 2051
QY 721 POWNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 2052 POWNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2104

RESULT 8
ADR94534
ID ADR94534 standard; protein; 637 AA.
XX ADR94534;
XX
XX
DT 16-DEC-2004 (first entry)
XX
XX
DE Novel S. pneumoniae protein sequence, SEQ ID 3169.
XX
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
KW bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
XX
PN US6800744-B1.
XX
PD 05-OCT-2004.
XX
XX 30-JUN-1998; 98US-00107433.
XX
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
XX WPI; 2004-697205/68.
DR N-PSDB; ADR91931.
XX
XX New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 3169; 151pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC

CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ Sequence 637 AA;
Query Match 75.7%; Score 3048; DB 8; Length 637;
Best Local Similarity 99.3%; Pred. No. 3.4e-168;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 185 LAFAGDMRLFVKNDQKAEIKIRPEKIKETKSEYVPYVSSYGNVIEGEGDLSKNKPN 244
Db 15 LAFAGDMRLFVKNDQKAEIKIRPEKIKETKSEYVPYVSSYGNVIEGEGDLSKNKPN 74
QY 245 LTKMESGKIYSDSEKQYLLKNDNIILRGYALKVTTYNPGKTDMLGEGVYSKEDIATQ 304
Db 75 LTKMESGKIYSDSEKQYLLKNDNIILRGYALKVTTYNPGKTDMLGEGVYSKEDIATQ 134
QY 305 KANPNLRALETTIYADSRNVEDGSRSTQVLSALDGFNIIRYQVFTFMNDKGEAIDK 364
Db 135 KANPNLRALETTIYADSRNVEDGSRSTQVLSALDGFNIIRYQVFTFMNDKGEAIDK 194
QY 365 GNLVTDSSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 424
Db 195 GNLVTDSSKLVLFQKDDKEYTGEDKSNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 254
QY 425 KIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLYLYGDLHDNTRDFNKLNVKDGDI 484
Db 255 KIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLYLYGDLHDNTRDFNKLNVKDGDI 314
QY 485 MDWGMKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEPVNIDPKGN 544
Db 315 MDWGMKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEPVNIDPKGN 374
QY 545 TSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSNDIKQIIDKTLNLIK 604
Db 375 TSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSNDIKQIIDKTLNLIK 434
QY 605 DFARNTTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKGE 664
Db 435 DFARNTTVKEFILNKDGTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKGE 494
QY 665 YQFDGWEISGPEGKXDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDN 724
Db 495 YQFDGWEISGPEGKXDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDN 554
QY 725 HSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 555 HSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 603
RESULT 9
AEA58404
ID AEA58404 standard; protein; 637 AA.
XX AEA58404;
XX


```
DE Streptococcus pneumoniae SP0043 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
FN WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019422.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
XX WPI; 1998-272224/24.
XX
XX N-PSDB; AAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX - or their epitope-containing fragments, useful in protective or
XX therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose
XX
XX Sequence 117 AA;
XX
XX Query Match 15.3%; Score 615; DB 2; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-28;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 657 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFKVPFKIEEKKKEENKPTFDVSK 716
XX Db 1 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFKVPFKIEEKKKEENKPTFDVSK 60
XX
XX QY 717 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 773
XX Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 13
XX ABP54590
XX ID ABP54590 standard; protein; 117 AA.
XX
XX AC ABP54590;
XX
XX DT 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
```

```
PN US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2002-479261/51.
XX
XX N-PSDB; ABQ84825.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX and for preventing or attenuating disease caused by Streptococcus
XX infection.
XX
XX Claim 11; Page 29; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX pneumoniae antigens have antibacterial activity and can be used in
XX vaccines. The S. pneumoniae antigens can also be used to prevent or
XX attenuate a Streptococcal infection in an animal. The polynucleotides
XX encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX of S. pneumoniae ORFs (open reading frames) which are used in an example
XX from the present invention
XX
XX Sequence 117 AA;
XX
XX Query Match 15.3%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-28;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 657 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFKVPFKIEEKKKEENKPTFDVSK 716
XX Db 1 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFKVPFKIEEKKKEENKPTFDVSK 60
XX
XX QY 717 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 773
XX Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 14
XX ADC45149
XX ID ADC45149 standard; protein; 117 AA.
XX
XX AC ADC45149;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX S. pneumoniae antigenic protein SP043.
XX
XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
XX Streptococcus pneumoniae.
XX
XX US6573082-B1.
XX
XX 03-JUN-2003.
XX
XX 28-MAR-2000; 2000US-00536784.
XX
```

PR	31-OCT-1996;	96US-0029960P.	PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PR	30-OCT-1997;	97US-00961083.	PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX			XX	
PA	(HUMA-)	HUMAN GENOME SCI INC.	DR	WPI: 2003-029926/02.
XX			DR	N-PSDB; ACA28663.
PI	Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;		XX	
PI	Rosen CA;		XX	
XX			XX	
XX	WPI: 2003-764574/72.		XX	
DR	N-PSDB; ADC45148.		XX	
XX			PS	Claim 25; SEQ ID NO 52737; 1766pp; English.
PT	Novel polynucleotide encoding Streptococcus pneumoniae polypeptides		XX	
PT	useful for producing vaccines for prevention or attenuation of infection		XX	
PT	by Streptococcus pneumoniae.		XX	
XX			XX	
PS	Example 1; SEQ ID NO 68; 58pp; English.		XX	
XX			XX	
CC	The invention relates to an isolated polynucleotide consisting of a		XX	
CC	Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding		XX	
CC	SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae		XX	
CC	antigens. Also included are making a recombinant vector by inserting the		XX	
CC	nucleic acid into a vector, an isolated polynucleotide consisting of at		XX	
CC	least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a		XX	
CC	recombinant host cell comprising the SP028 polynucleotide. The nucleic		XX	
CC	acids are useful as DNA vaccine against Streptococcus pneumoniae		XX	
CC	infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae		XX	
CC	antigen nucleic acids are useful as probes for use in diagnostic methods		XX	
CC	for detecting S. pneumoniae gene expression. The present sequence		XX	
CC	represents an S. pneumoniae antigenic protein.		XX	
XX			XX	
SQ	Sequence 117 AA;		SQ	
Query Match			Query Match	
Best Local Similarity 15.3%; Score 615; DB 7; Length 117;			5.6%; Score 227; DB 6; Length 1196;	
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			Best Local Similarity 20.2%; Pred. No. 0.00025;	
			Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;	
QY	657 YKGELEKGYQFDGWEISGPEGKADAGYVNLKSDTFIKPVFKKIEKKEENKPTFDVSK 716		QY	1 KLGEIAESKFKNLGNKGEG-----SLKDDTTGVHHHQBNEESI 39
Db			Db	123 RLKDIQE-LFMDTGIGKEGYIIGGKIEAVLSGKPEERRSLLEAAGIVFKWKREAD 181
QY	717 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 773		QY	40 KEKSFITDRNISTIRDPEN--KDLKKLIKKKFREVDFF--TSETGKME-----85
Db			Db	182 KKLNS--TEONLIRIKDILNTVEERMEPLKESEKAKKFLNLSEELKKEVNVMTYSIDK 239
QY	61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117		QY	86 -EYDYKYDDKGNITAYDDGTD-LEYETKLEDEIKSIYGVLSPSKDGHEIFILGKISNVSK 143
Db			Db	240 TEKDLK-NISSNMLSLGENIDNLNKEKSQYKEIISKF-----NEKLELDK--NNSR 288
QY			QY	144 NAKVYVGNYSIEIKATKYDFHSKTMFTFDLYANINDIVDGLAFAGDMRLFVKDNDQKA 203
Db			Db	289 NKEEYVNDKNDKDI-----NENALLKEKIKNLKD-----NIKVENTLKTN 331
QY			QY	204 EIKIR-----MPEKIKETKSEYPVY-----223
Db			Db	332 EERLQKYLKEKEALEKKINKLKEESYLRBEIGKEDNVNFKELKEKEKLTJKSEE 391
QY			QY	224 -----SSYGNVIELGEGDL--SKNRPDNTUKMESGKIYS-----DSEKQOYLK 265
Db			Db	392 TEILSNTSLNRNEISVMENEIVNLKNDIKNSCDSYISSININIKTKDIEIKNIK 451

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:49:07 ; Search time 181 Seconds
(without alignments)
1978.260 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAESKFKNLGNKGEKS.....ATVLKNNISSKSTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_Main:*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	773	4	US-10-067-385-8
2	4026	100.0	2119	3	US-09-769-744A-28
3	4026	100.0	2140	4	US-10-282-122A-73670
4	4026	100.0	2140	5	US-10-472-928-1180
5	3048	75.7	637	5	US-10-617-320-3169
6	750.5	18.6	1529	5	US-10-617-320-4771
7	615	15.3	117	3	US-09-765-272-68
8	615	15.3	117	6	US-11-106-649-68
9	230	5.7	1404	5	US-10-732-923-3304
10	227	5.6	1196	4	US-10-282-122A-52737
11	222	5.5	861	3	US-09-820-843A-109
12	221.5	5.5	1441	5	US-10-732-923-3352
13	218	5.4	1602	5	US-10-954-924-7
14	218	5.4	1621	5	US-10-954-924-5
15	218	5.4	1639	4	US-10-087-464-10
16	218	5.4	1639	5	US-10-954-924-3
17	216	5.4	2399	5	US-10-732-923-15036
18	212	5.3	1141	4	US-10-282-122A-70251
19	208	5.2	2310	5	US-10-732-923-3342
20	207.5	5.2	4688	4	US-10-282-122A-76865
21	207	5.1	1191	4	US-10-282-122A-52048
22	204.5	5.1	1434	5	US-10-732-923-4235
23	204	5.1	1166	5	US-10-470-048B-153
24	204	5.1	1419	5	US-10-732-923-4285
25	203.5	5.1	1985	5	US-10-732-923-3351
26	203	5.0	1166	5	US-10-744-672-7
27	203	5.0	1166	5	US-10-744-616-7

ALIGNMENTS

RESULT 1

US-10-067-385-8
; Sequence 8, Application US/10067385
; Publication No. US20020110562A1
; GENERAL INFORMATION:
; APPLICANT: Adamou, John
; APPLICANT: Choi, Gil
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
; FILE REFERENCE: 469201-589
; CURRENT APPLICATION NUMBER: US/10/067,385
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US/09/590,991
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US/60/138,453
; PRIOR FILING DATE: 1999-06-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 773
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-067-385-8

Query Match 100.0%; Score 4026; DB 4; Length 773;
Best Local Similarity 100.0%; Pred. No. 2.1e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	KLGETAESKFKNLGNKGEKSLLKDTTGVVHHHQNNEESIKEKSSFTIDRNISTIRDFENK	60
Db	1	KLGETAESKFKNLGNKGEKSLLKDTTGVVHHHQNNEESIKEKSSFTIDRNISTIRDFENK	60
Qy	61	DLKKLKKKFFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLDIKSKI	120
Db	61	DLKKLKKKFFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLDIKSKI	120
Qy	121	YGVLSPSKDGHEIFILGKISNVSKNAKYVYNNYKSIIEIKATKYPHSHKTMTPDLYANIND	180
Db	121	YGVLSPSKDGHEIFILGKISNVSKNAKYVYNNYKSIIEIKATKYPHSHKTMTPDLYANIND	180
Qy	181	IVDGLAFAGDMRLPFVKNDQKAEIKIRMPKIKETKSEYPVVSSYGVNVIELGEGDLSKN	240
Db	181	IVDGLAFAGDMRLPFVKNDQKAEIKIRMPKIKETKSEYPVVSSYGVNVIELGEGDLSKN	240
Qy	241	KPDNLTKMESKIIYSDSEKQQYLLKDNIIILKGLYALKVTTYNPGKTDMLGNGVYSKEDI	300
Db	241	KPDNLTKMESKIIYSDSEKQQYLLKDNIIILKGLYALKVTTYNPGKTDMLGNGVYSKEDI	300
Qy	301	AKIQKANPNLRALSETTIYADSRNVGDSRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA	360
Db	301	AKIQKANPNLRALSETTIYADSRNVGDSRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA	360

Qy 361 IDKGNLVTSSKLVLFQKDDKEYTGBDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 361 IDKGNLVTSSKLVLFQKDDKEYTGBDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFWNLRVNESVDNYLYGDLHIDNTRDFNKLNVK 480
Db 421 SKSNKIYVRNPEFYLRGKISDKGFWNLRVNESVDNYLYGDLHIDNTRDFNKLNVK 480
Qy 481 DGDIMDMGMDYKANGFPDKVTMDGNVYLQTSYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db 481 DGDIMDMGMDYKANGFPDKVTMDGNVYLQTSYSDLNKAVGVHYQFLYDNVKEPVNID 540
Qy 541 PKGNTSIEYADGKSVFVNINDKRNNGFDEGEIOQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 541 PKGNTSIEYADGKSVFVNINDKRNNGFDEGEIOQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
Qy 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Db 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Qy 661 LEKGYQFDGWEISGFEGKDAAGYVNLNLSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 720
Db 661 LEKGYQFDGWEISGFEGKDAAGYVNLNLSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 720
Qy 721 PQVHSQLNESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 773
Db 721 PQVHSQLNESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 773

RESULT 2

US-09-769-744A-28
; Sequence 28, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard Wf
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hanebro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match 100.0%; Score 4026; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 7.6e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KLGEIAESKFNKLGNGKGGSLKKTGTGVEHHQHNEESIKEKSSFTTDRNISTIRDPENK 60
Db 1313 KLGEIAESKFNKLGNGKGGSLKKTGTGVEHHQHNEESIKEKSSFTTDRNISTIRDPENK 1372
Qy 61 DLKLIKKKFREVDFTSETCKRMEEDYKYDDKGNIIAYDDGTDLEVEKLEIKSKI 120
Db 1373 DLKLIKKKFREVDFTSETCKRMEEDYKYDDKGNIIAYDDGTDLEVEKLEIKSKI 1432
Qy 121 YGVLSPSKDGHFELGKISNVSNAKVVYGNYSIEIKATKYDFHSKTMTFDLYANIND 180
Db 1433 YGVLSPSKDGHFELGKISNVSNAKVVYGNYSIEIKATKYDFHSKTMTFDLYANIND 1492

Qy 181 IVDGLAPAGMRFLFVKNDQKAEIKIRMPBEKIKETHKSEYPYVSSYGNVIELGEGDLSKN 240
Db 1493 IVDGLAPAGMRFLFVKNDQKAEIKIRMPBEKIKETHKSEYPYVSSYGNVIELGEGDLSKN 1552
Qy 241 KPNLNTWESGKLYSDEKQOYLLKKNIIIRKGVALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1553 KPNLNTWESGKLYSDEKQOYLLKKNIIIRKGVALKVTTYNPGKTDMLGNGVYSKEDI 1612
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 360
Db 1613 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 1672
Qy 361 IDKGNLVTSSKLVLFQKDDKEYTGBDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1673 IDKGNLVTSSKLVLFQKDDKEYTGBDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 1732
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFWNLRVNESVDNYLYGDLHIDNTRDFNKLNVK 480
Db 1733 SKSNKIYVRNPEFYLRGKISDKGFWNLRVNESVDNYLYGDLHIDNTRDFNKLNVK 1792
Qy 481 DGDIMDMGMDYKANGFPDKVTMDGNVYLQTSYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db 1793 DGDIMDMGMDYKANGFPDKVTMDGNVYLQTSYSDLNKAVGVHYQFLYDNVKEPVNID 1852
Qy 541 PKGNTSIEYADGKSVFVNINDKRNNGFDEGEIOQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1853 PKGNTSIEYADGKSVFVNINDKRNNGFDEGEIOQHIIYINGKEYTSFNDIKQIIDKTLNIK 1912
Qy 601 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Db 1913 IVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 1972
Qy 661 LEKGYQFDGWEISGFEGKDAAGYVNLNLSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 720
Db 1973 LEKGYQFDGWEISGFEGKDAAGYVNLNLSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 2032
Qy 721 PQVHSQLNESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 773
Db 2033 PQVHSQLNESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2085

RESULT 3

US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578


```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match          100.0%; Score 4026; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 7.7e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNKLGNGKESLKKDTTGVVHHHQENEESIKKSSFTIDRNISTIRDFENK 60
   |||||
Db 1334 KLGEIAESKFNKLGNGKESLKKDTTGVVHHHQENEESIKKSSFTIDRNISTIRDFENK 1393

QY 61 DLKLLIKKKFREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLLEYETEKLEIKSKI 120
   |||||
Db 1394 DLKLLIKKKFREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLLEYETEKLEIKSKI 1453

QY 121 YGVLSKDGHEIFLIGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSTKMTFDLYANIND 180
   |||||
Db 1454 YGVLSKDGHEIFLIGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSTKMTFDLYANIND 1513

QY 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPEIKETKSEYPVYSSGVNVIELGEGDLSKN 240
   |||||
Db 1514 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPEIKETKSEYPVYSSGVNVIELGEGDLSKN 1573

QY 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
   |||||
Db 1574 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633

QY 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFKMNDKGEA 360
   |||||
Db 1634 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFKMNDKGEA 1693

QY 361 IDKGNLVTDSSKLVLFKGDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
   |||||
Db 1694 IDKGNLVTDSSKLVLFKGDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1753

QY 421 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 480
   |||||
Db 1754 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 1813

QY 481 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLQYGYSDLNKAVGVHYQFLYDNVKEPVNID 540
   |||||
Db 1814 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLQYGYSDLNKAVGVHYQFLYDNVKEPVNID 1873

QY 541 PKGNTSIEYADGKSVFVNDKRNNGPDGEIOEQHIYINGKEYTSFNDIKQIIDTLNIK 600
   |||||
Db 1874 PKGNTSIEYADGKSVFVNDKRNNGPDGEIOEQHIYINGKEYTSFNDIKQIIDTLNIK 1933

QY 601 IVVKDFARNTTVKFEILNKDTEGEVSELKPHRVTVVTIQNGKEMSSSTIVSEEDFILPVYKGE 660
   |||||
Db 1934 IVVKDFARNTTVKFEILNKDTEGEVSELKPHRVTVVTIQNGKEMSSSTIVSEEDFILPVYKGE 1993

QY 661 LEKGYQFDGWEISGPEGKADAGVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKXDN 720
   |||||
Db 1994 LEKGYQFDGWEISGPEGKADAGVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKXDN 2053

QY 721 PQVNHSQLNESHKEDLQREEHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
   |||||
Db 2054 PQVNHSQLNESHKEDLQREEHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106
```

RESULT 4

```

US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication NO. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180
```

```

Query Match          100.0%; Score 4026; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 7.7e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNKLGNGKESLKKDTTGVVHHHQENEESIKKSSFTIDRNISTIRDFENK 60
   |||||
Db 1334 KLGEIAESKFNKLGNGKESLKKDTTGVVHHHQENEESIKKSSFTIDRNISTIRDFENK 1393

QY 61 DLKLLIKKKFREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLLEYETEKLEIKSKI 120
   |||||
Db 1394 DLKLLIKKKFREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLLEYETEKLEIKSKI 1453

QY 121 YGVLSKDGHEIFLIGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSTKMTFDLYANIND 180
   |||||
Db 1454 YGVLSKDGHEIFLIGKISNVSKNAKVYGNKYKSIIEIKATKYDFHSTKMTFDLYANIND 1513

QY 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPEIKETKSEYPVYSSGVNVIELGEGDLSKN 240
   |||||
Db 1514 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPEIKETKSEYPVYSSGVNVIELGEGDLSKN 1573

QY 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
   |||||
Db 1574 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633

QY 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFKMNDKGEA 360
   |||||
Db 1634 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFKMNDKGEA 1693

QY 361 IDKGNLVTDSSKLVLFKGDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
   |||||
Db 1694 IDKGNLVTDSSKLVLFKGDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1753

QY 421 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 480
   |||||
Db 1754 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDHLHIDNTRDFNKLNVK 1813

QY 481 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLQYGYSDLNKAVGVHYQFLYDNVKEPVNID 540
   |||||
Db 1814 DGDIMDWGMKDYKANGFPDKVTDMDGNVYLQYGYSDLNKAVGVHYQFLYDNVKEPVNID 1873

QY 541 PKGNTSIEYADGKSVFVNDKRNNGPDGEIOEQHIYINGKEYTSFNDIKQIIDTLNIK 600
   |||||
Db 1874 PKGNTSIEYADGKSVFVNDKRNNGPDGEIOEQHIYINGKEYTSFNDIKQIIDTLNIK 1933

QY 601 IVVKDFARNTTVKFEILNKDTEGEVSELKPHRVTVVTIQNGKEMSSSTIVSEEDFILPVYKGE 660
   |||||
```

Db 1934 I V V K D F A R N T T V K G F I L N K D T G E V S E L K P H R V T V T I O N G K E M S S T I V S E E D F I L P V Y K G E 1993

Qy 661 L E K Y Q D G W E I S G F E G K K D A G Y V I N L S K D T F I K P V F K K I E E K E E B N K P T F D V S K K K D N 720

Db 1994 L E K Y Q D G W E I S G F E G K K D A G Y V I N L S K D T F I K P V F K K I E E K E E B N K P T F D V S K K K D N 2053

Qy 721 P O V N H S Q L N E S H R K E D L Q R E H S Q K S D S T K D V T A T V L D K N N I S S K S T T N N P N K 773

Db 2054 P O V N H S Q L N E S H R K E D L Q R E H S Q K S D S T K D V T A T V L D K N N I S S K S T T N N P N K 2106

RESULT 5

US-10-617-320-3169

; Sequence 3169, Application US/10617320

; Publication No. US20050136404A1

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

; THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660

; COMPUTER: <Unknown>

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: <Unknown>

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/617,320

; FILING DATE: 10-Jul-2003

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,433

; FILING DATE: 30-Jun-1998

; APPLICATION NUMBER: 60/ 085131

; FILING DATE: May 12, 1998

; APPLICATION NUMBER: 60/051553

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-011

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 3169:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 637 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus pneumoniae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (B) LOCATION 1...637

; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:

US-10-617-320-3169

Query Match 75.7%; Score 3048; DB 5; Length 637;

Best Local Similarity 99.3%; Pred. No. 6.5e-162;

Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 185 L A F A G D M R L F V K D N D Q K A E I K I R M P E K I K T K S E Y P V S S Y G N V I E L G E G D L S K N K P D N 244

Db 15 L A F A G D M R L F V K D N D Q K A E I K I R M P E K I K T K S E Y P V S S Y G N V I E L G E G D L S K N K P D N 74

Qy 245 L T K W E S G K I Y S D S E K Q O Y L L K D N I I L R K G Y A L K V T T Y N P G K T D M L E G N G V Y S K E D I A K I Q 304

Db 75 L T K W E S G K I Y S D S E K Q O Y L L K D N I I L R K G Y A L K V T T Y N P G K T D M L E G N G V Y S K E D I A K I Q 134

Qy 305 K A N P N L R A L S E T T I Y A D S R N V E D G R S T Q S V L M S A L D G F N I I R Y Q V F T F K M N D K G E A I D K D 364

Db 135 K A N P N L R A L S E T T I Y A D S R N V E D G R S T Q A V L M S A L D G F N I I R Y Q V F T F K M N D K G E A I D K D 194

Qy 365 G N L V T D S S K L V L F G K D D K E Y T G B D K F N V E A I K E D G S M L F I D T T K P V N L S M D K N Y F N P S K S N 424

Db 195 G N L V T D S S K L V L F G K D D K E Y T G B D K F N V E A I K E D G S M L F I D T T K P V N L S M D K N Y F N P S K S N 254

Qy 425 K I Y V R N P E F Y L R G K I S D K G G F N W E L F V N E S V V D N Y L I Y G D L H I D N T R D F N I K L N V K D G D I 484

Db 255 K I Y V R N P E F Y L R G K I S D K G G F N W E L F V N E S V V D N Y L I Y G D L H I D N T R D F N I K L N V K D G D I 314

Qy 485 M D W G M K D Y K A N G F P D K V T D M D G N V I L Q T G Y S D L N A K A V G V H Y Q F L Y D N V K P E V N I D P K G N 544

Db 315 M D W G M K D Y K A N G F P D K V T D M D G N V I L Q T G Y S D L N A K A V G V H Y Q F L Y D N V K P E V N I D P K G N 374

Qy 545 T S I E Y A D G K S V V F N I N D K R N N G F D G E I Q E O H I Y N G E Y T S F N D I K O I I D K T L N I K I V W K 604

Db 375 T S I E Y A D G K S V V F N I N D K R N N G F D G E I Q E O H I Y N G E Y T S F N D I K O I I D K T L N I K I V W K 434

Qy 605 D F A R N T T V K F I L N K D T G E V S E L K P H R V T V T I O N G K E M S S T I V S E E D F I L P V Y K G E L E K G 664

Db 435 D F A R N T T V K F I L N K D T G E V S E L K P H R V T V T I O N G K E M S S T I V S E E D F I L P V Y K G E L E K G 494

Qy 665 Y Q F D G W E I S G F E G K K D A G Y V I N L S K D T F I K P V F K K I E E K E E B N K P T F D V S K K K D N P Q V N 724

Db 495 Y Q F D G W E I S G F E G K K D A G Y V I N L S K D T F I K P V F K K I E E K E E B N K P T F D V S K K K D N P Q V N 554

Qy 725 H S Q L N E S H R K E D L Q R E H S Q K S D S T K D V T A T V L D K N N I S S K S T T N N P N K 773

Db 555 H S Q L N E S H R K E D L Q R E H S Q K S D S T K D V T A T V L D K N N I S S K S T T N N P N K 603

RESULT 6

US-10-617-320-4771

; Sequence 4771, Application US/10617320

; Publication No. US20050136404A1

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

; THERAPEUTICS

; NUMBER OF SEQUENCES: 5206

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660

; COMPUTER: <Unknown>

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: <Unknown>

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/617,320

; FILING DATE: 10-Jul-2003

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,433

; FILING DATE: 30-Jun-1998

; APPLICATION NUMBER: 60/ 085131

; FILING DATE: May 12, 1998

; APPLICATION NUMBER: 60/051553

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-011

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 3169:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 637 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus pneumoniae

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (B) LOCATION 1...637

; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:

US-10-617-320-4771

Query Match 75.7%; Score 3048; DB 5; Length 637;

Best Local Similarity 99.3%; Pred. No. 6.5e-162;

Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 185 L A F A G D M R L F V K D N D Q K A E I K I R M P E K I K T K S E Y P V S S Y G N V I E L G E G D L S K N K P D N 244

Db 15 L A F A G D M R L F V K D N D Q K A E I K I R M P E K I K T K S E Y P V S S Y G N V I E L G E G D L S K N K P D N 74

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4771:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1529
; SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-10-617-320-4771

Query Match      18.6%; Score 750.5; DB 5; Length 1529;
Best Local Similarity 79.7%; Pred. No. 3.6e-33;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY 1 KLGETAESKFKNLGKGLKKTGTGVHHQENEEISKEKSSFTIDRNISTIRDFENK 60
Db 1336 KLGETAESKFKNLKVKDLSLNKETAENVENLLVONQSLFNHKTISTIRDFENK 1395

QY 61 DLKLIKKKFKREVDFTSETGKRMMEYDYKDDKGNIIAYDDGTDLLEYETEKLDIKSKI 120
Db 1396 DLKLIKKKYKQEDFVN-GGRTTVERDYKDDKGNIIAYDDGTDLLEYETEKLDIKSKI 1454

QY 121 YGVLSPKDGHFEILKIGNSVKNKAVYVGNKYKSIETKATKYDFHSTKMTTDPDLYANIND 180
Db 1455 YGVLSPKDGHFEILKIGNSVKNKAVYVGNKYKSIETKATKYDFHSTKMTTDPDLYANIND 1514

QY 181 IVDGLAF 187
Db 1515 IVDGISF 1521

RESULT 7
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061549A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512

```

```

; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match      15.3%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 5e-27;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPK 773
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPK 117

RESULT 8
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      15.3%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 5e-27;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKIIEKKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPK 773
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPK 117

RESULT 9
US-10-732-923-3304
; Sequence 3304, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154

```



```
QY 86 -EVDYKDDKGNIIAYDDGTD-LEYETEKLEIKSIYGVLSPSKDGHPFELGKISNVSK 143
Db 240 IEKDLK-NISSNLSLGENIDLNKESQYKEIISKF-----NEKLELLDK--NNSR 288
QY 144 NAKVYGNKYKSEIKATKYDPHSKTTWTFDLVANINDIVDGLAFAGDMRLFVKNDQKKA 203
Db 289 NKEEYNNKDKNDIE-----NENALLKEKIKNLKD-----NIKVKENTLKTN 331
QY 204 EIKIR-----MPEKIKETSEYPYV----- 223
Db 332 EERLOKYLKEALEKKINKLKEESYLREBEICKEDNVNPNKELKEKEKLIKLSSEE 391
QY 224 -----SSYGNVIELGEGDL--SKNPNLNTKMSGKIYS-----DSEKQOYLK 265
Db 392 EIBLSNTSLRNEISWENEIVLENKLDNIKNKSCDSYISSINIKTKEDIEKEIKNIK 451
QY 266 DNILL-----RKGYALKVTTYNPCKTMDLEGNGVYSKED-----IAKTQK----AN 307
Db 452 ENILLLENLKNRSNMISSLKISLNNKEK-KLKEKNAAYSRLLEANVHMLSNLEKHYEGYN 510
QY 308 PNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKGBAIDKGNL 367
Db 511 RSVKTLMEHVSQKVDNIKGG-----CEVLGDIIVKVKKELETAMEIALGGAI---SNV 560
QY 368 VTDSSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP---SKSN 424
Db 561 ITEDE-----NKAKILINYLKKK-----SLGRATFLPLTTIQGR 594
QY 425 KIYVRNPEFLRGKISDKGFPWELRVNESVVDNYLIYGDLDHIDNTRDNI---KLNVDKG 482
Db 595 KAKINN-----VTREDGF---LGIASDLIDY-----DVKFSNIIIDYVLGRTLVAKD- 637
QY 483 DTMWGMKDYKANGFPDKVTDMDGNVYLQTVGSYDLNAKAVGHYQFLYDNVKNPEVNIDPK 542
Db 638 ---MDSALKIATKLNYSFKLVILEGEV-----INPG 665
QY 543 GNT-----SIEYADGKSVFNINDKRNNGDFGEIQEHYINGKEYTSFNDIKIIDKTLNI 599
Db 666 GSLTGSIKHRAAGSSI---ISKR-----EIEE-----TKKELEBETKN- 700
QY 600 KIVVKDFARNITTVKFEILNKDTGVESELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 659
Db 701 -----FIEBFMGN-----ILENKNKIKT--LDEENLNI---KD 728
QY 660 ELEKGYQPDGWEISGFEGKDA-----GYVINLSKD--TFIKPVFKIIEKKEBENK 709
Db 729 EI-----YNNIEITKFMGKLNAIKEDTELRSSLNISREIEIKLTDKDKIDENINESQK 784
QY 710 PTFDVSKKKDNQVNHSQLNESHKREDLQREHSQKSDTK-----DVTATVLDKNNISSK 765
Db 785 QLEELKLLKD---LNHNDIKEC--BDFLQNEEVENKNIKDKLIEYKIEKAKLDENMLVSIK 839
QY 766 -----STNNPNK 773
Db 840 KELYSMDTNITNLNENK 857

RESULT 11
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
; LENGTH: 861
```

```
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|4493994
US-09-820-843A-109

Query Match
Best Local Similarity 5.5%; Score 222; DB 3; Length 861;
Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

QY 2 LGEIAESKFKNLGNG-----KEGSILKDDTTGVVEHHH--QENEESIEKSSFTI 47
Db 154 LNNINGKFKNLDDCYNLPTNNLIYDKEG--KHLTGKEHYNAASSNEYNNKNNKNNY 211
QY 48 DRNISTIRDFENKOL-----KKLIKPKPREVDDPTSETGRMEEEYDYKDDKGNIT 98
Db 212 NNNSYNNNNFNNYNDNNYNNNNKMGKNGKYSRLNVLK---KEHDMVDYEYNNKGNIR 268
QY 99 AYD-----DGTDLVEYETEKLEIKSIYGVLSPSKDGHPFI--LGKISNVSKNAKYVGN 151
Db 269 KNDSEKYWDNPLHY-----SKQNNYDIFTLGDIKKYAKNKECKGN 310
QY 152 NYKSEIKATKYDPHFSKTTMTFDLYANINDIVDG---LAFAGDMRLFVKNDOKKAEIKIR 208
Db 311 KYMNM-----HDNNSN-----NSNNVLNNNNNSNNYNNIFKNDDENL----- 351
QY 209 MPEKIKETKSEYPVSSYGNVIELGEG-DLSK---NKPDLNTKMSGKIYSDSEKQOYL 264
Db 352 -----TKSNFAKWFKNNNNNVNNENTDIIKLANK-----NSQHSQDKGNNNNN 396
QY 265 KDNILRGYALKVTTYNPCKTMDLEGNGVYSKEDIKIQKANPN--LRALSETTIYAD 321
Db 397 GNNII-----NNNSNNKNIPQGNRSRYENVMYNNNNNNNNIISNNKNEASFND 447
QY 322 SRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKG--EADKDGMLVTDSSKLVLFGK 380
Db 448 NINTNSGREEEKISNTVAE---LLMKQISMLKERNKGLDVLKQNT-----FGFL 494
QY 381 DKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNPSKNIYVRNPEFYL----- 435
Db 495 DNNYQNYGSNN-----NSSLEKNNM---KENDIYSKEASKRIMDI PR 533
QY 436 ---RGKISDKGFPWELRVNESVVDNYLIYGDLDHIDNTRDNIKLVKDGDMWGMKD 491
Db 534 TLNNGLVSQE-----SLLVNQSVLNNNNNNYNNNNNNNNNNNNNNNNNNNNNNM----- 582
QY 492 YKANGFPDKVTDMDGNVYLQTVGSYDLNAK---AVGVHYQFLYDNVKNPEVNIDPKGNTSTE 548
Db 593 -----NNMNSNNNNNNNNNNYNNNNYNNNNYNNNNNNNNNNNNNNNNNNNNNNNN 633
QY 549 YADGKSVFNINDKRNNGDFGEIQEHYINGKEYTSFNDIKQIIDK----- 595
Db 634 NSNNVHEHYNNKKN--FKNKINNYHNLDPNKNMMNNNNNTYNNINKNNLSNMENFPFSL 691
QY 596 TLNLIKIVVVKFARN---TTVKEFTLANKDTGEVSELKPHRVTVTIQNGKMSSTIVSEED 651
Db 692 SFNNSDINKNNAQGNINITPTIINSILRLD-NEVDNVMHNSISENTQNAK--VSNVLSLK 748
QY 652 FILPVYKGELEKGYQPDGWEISGFEGKDAQGVNLSKDTFIK-----PVFKKIEEKK 704
Db 749 SLLKASKSQGNNNYNIPI-----KFNNNNN-----NNNSKFINYNSQOYFSHQOQOQH 799
QY 705 EEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSQK---SDSTKDVATVATVLDK 759
Db 800 QOQOQOQ--QOQOQTLIQTINGTHLNDFNKKKFNKKERYPMKYPEPDGTTTNETMMVREK 855

RESULT 12
US-10-732-923-3352
; Sequence 3352, Application US/10732923
```

```
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3352
; LENGTH: 1441
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; US-10-732-923-3352

Query Match      5.4%; Score 221.5; DB 5; Length 1441;
Best Local Similarity 19.5%; Pred. No. 0.0011;
Matches 168; Conservative 164; Mismatches 286; Indels 245; Gaps 48;

Qy 4 EIAESKFNUG-----NGKE-GSLKXD--TTGVEHHH-QENESIKEKSF-----T 46
Db 84 ELLKEKELINIKENKINNKEIITLKKEELNDIENEYIEKNKE--KEKLNYEITNIKMS 141
Qy 47 IDRNISTIRDPENKDLKLKK-----KPEVDDFTSETCKRMEEDYKYDDKGNLIAY 100
Db 142 LDKLSCSIQE--KENLEKINKKVVKEKNRELKEFKMEKNEIIESLNKTIIDDKN--AY 198
Qy 101 DD-GTDLEYETKDETKSKYIGVLSPSKDGHF-----ILGKISNV----- 142
Db 199 EKLETFEKKEMLEMDSKLI-----EKEKKFIKKEKLEKENEVMEKLDIENKEEH 253
Qy 143 -KNKAVYGNKYSI-----BIKATYDFHSTKTTFDLY-ANIINDIVDGLAFAGDMRLFV 195
Db 254 FKNKEEFPKKEEKFENLENELNKLKSLDSKNACQMEIYKMEIKDLSQSLV----- 304
Qy 196 KDNOKKAEIKIRPEKIKETKSEVPYVSSYG-----NVELGEGDLSKNKPDNLTKME 249
Db 305 -EKEREIFEIKNEVDDKINNKNK--LSNFNNDNDNTVKCSEELINKKIEAVNKL- 359
Qy 250 SGKIYSDSEKQOYLKONILIRKGYALKVTTYPNPGKTMLEGNVGYSGEDIAKIQKANPN 309
Db 360 -----IKEKEMELNE--IHKKYNLEIEKIK-----NEINEKEKELEQNKKHN 400
Qy 310 LRALSETT--IYADSRNVDEGRSTQSVLMSALDG-----FNIRYQVFTF 352
Db 401 IE-INDLTKEIQIREKKIEDYKEEYKIELSKLDSKKNKIENNELNNEVNSLNNEVNSL 459
Qy 353 -----KWNKGEAIDKGNLYTDSKLV--LFGKDDKEYTGEDKFVNEAIKED----- 398
Db 460 NNEVNSLNEINSLNNDKQTLUSKNKLLNDLNNLKNEINNSDN--KWNKMKEDIIMLNEE 518
Qy 399 --GSMLEFDTKPNVLSMDKNYFNPSKSNKIYVRNPEFYLRGKIDGKGFNWLVRNBSV 456
Db 519 LEGKCVVD-----EIEKKY-----KNEIFI-----FEKKLAKEN-- 549
Qy 457 DNYLIYDGLHDITRDNFNKLVNKGDIIMGMKMDYKANGPPDKVTMDGVNLYQTGYSD 516
Db 550 -----YDDLN--DEINILANDISVKEKEFIE--MKEF-----YENKINLYKNFEEKTKYYE 597
Qy 517 LNAKAVGHVYQFLVDNYK--PEVNIDPKGNTSIEYAD-----GKSVFNINDKENNG 566
Db 598 NEMILIRUKYNNEOQLIKQIDELNIQ--KLKTEEKYQLYDNDNIKFKSICAKINTTYSN 656
Qy 567 FDGE--IOEQHIYINGKEYTSFNDIKQIIDKTLNKIIVVKDFARNTTVKPEFILNKDTGCV 624
Db 657 IKGSDLVNFVEAYIKHRDESFD-----ANESGIKEVVAE-----LEEKHREEV 701
Qy 625 SELKPHRTVTVIQNGKEMSSIV--SEEDFTLPVYKGELEKGYFDGWEISGFEGKADAG 682
Db 702 SRIREEHKKYVRIEGEEHKEEVRIGEE-----HKEEVK-----IGEQRBELG 746
Qy 683 YVINLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNPQVNHSQLNESHREKDLQ--REE 741

; Publication No. US2005010954924
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; APPLICANT: PAN, WEIQING
; TITLE OF INVENTION: Recombinant Process for Preparing a Complete Malaria Antigen GP19
; FILE REFERENCE: GRUE-003DIV
; CURRENT APPLICATION NUMBER: US/10/954,924
; CURRENT FILING DATE: 2004-09-29
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Versior. 4.0
; SEQ ID NO 7
; LENGTH: 1602
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-10-954-924-7

Query Match      5.4%; Score 218; DB 5; Length 1602;
Best Local Similarity 18.6%; Pred. No. 0.002;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VEHHHQENESIK-----FKSFTIDRNISTIRDPENKDLKLKKKRFREVD 75
Db 213 KDNVGMEDYIKNNKKTINENILIEESKKTIDKNKNTATKEE-----KKLYQAOY 264
Qy 76 FTSETGRMEEDYKYDDKGNLIAYDDGTDLEYETKDEIKSKIYGVLSPSKDGHPFIL 135
Db 265 DLSIYNKQLEE-----AHNLISV-----LE---KRIDTLK-----KNENIKELL 300
Qy 136 GKISNVSKNAKVYGNKYSIKAIKYDPH-----SKTMTFDLYANIINDIVDGLAFA 188
Db 301 DKINEIKNPPANSNGTPTNLLDKNKKIEBEKEIKEIAKTIKFNIDSLFTDPL----- 354
Qy 189 GDMRLFPKNDQKAEIKIRMPKIKETKSEYP-----YVSSYGNV-----TELGE-GDLSK 239
Db 355 -ELEYLREKN-KNIDISAKVETKSETPENYNGVTPYPLSYNDINNALNELNSFGDLI- 411
Qy 240 NKPDNLTKMESGKIYSDSEKQOYL--LKDNIL-----RKGYALKVTTYPNPGKTDML 289
Db 412 -NPFYTKESKNYITDNERKKFINIEKIEKIEKESDKKSYEDRSKSLN-----DIT 466
Qy 290 EG-----NGVYSKE-----DIAKIQKA-----NPNLHALSETTIYADSRNVDEGRSTQSV 334
Db 467 KEYEKLLNEIYDSKFNNNIDLTNFKMGKRYSYKVVKLTHHNTTFASYENSKNHLEKLT 526
Qy 335 LMSALDGF--NII-----RYQVFFPKMNDKGEAIDKGNLYTDSKLVLFKDKDKKEYT 385
Db 527 ALKYMEDYSURNIVWEKELKYKNLASKIENEIETLVE--NIKKDEEQLF-----EKKIT 579
Qy 386 GEKDFNVEALKEGSMLEFIDTPVNLSDMKNYFNPSKSNKIYVRNPEFYLRGKISDKGGF 445
Db 580 KDNKPKPEKILEVSDIVKVVQVKVLL---NNKIDELAKTQLILKQVE--LKGNIHVNSY 634
Qy 446 NWELRVNESVVDNYLYGDLHDITRDNFNKL-----NVK----- 480
Db 635 KQENKQE---PYVLIIVLKEIKDLKVFMPKPVESLINEEKKNIKTEGQSONSEPTGEI 690
```


Db 232 KDNVGMEDYKKNKKTINENINELIEBSKKTIDKNKNAKTEE-----KKKLYQAQY 283
Qy 76 FTSETGRMEBYDYKDDKGNITAYDDGTGLEYTEKLEIKSKIYCVLSPSKDGHFEIL 135
Db 284 DLSYNRKQLEE-----AHNLISV-----LE---KRIDLK-----KNENIKELL 319
Qy 136 GKISNVSKNAKVYVYNNVKSIEIKATKYDPH-----SKTMTFDLYANINDIVDGLAPA 188
Db 320 DKINEIKNPPANGNTPTNLLDNKKIEHEKEIKIATIRFNIDSLFTDPL----- 373
Qy 189 GDMRLFVKNDQKAEIKIRMEKIKETKSEYP-----YVSSYGNV-----IELGE-GDLSK 239
Db 374 -ELEYLREKN-KNIDISAKVETKESTEPNEYPNGVTYPLSYNDINNALNELASFGDLI- 430
Qy 240 NKPDNLTKMESGKIYSSEKQYVL--LKDNIIL-----RKYALKVTTYNPCKTMDML 289
Db 431 -NPFDTKPEFSKNIYTDNERKKTINEIKIEKIEKIESDKSYEDRSKSLN-----DIT 485
Qy 290 EG-----NGVYSKE-----DIAKIQA-----NPNLRALSETTIYADSRNVEDGRSTQSV 334
Db 486 KEYEKLNEIYDSKFNNDILDTNFKMMGKRYSYKVEKLTHNTTFASYENSKNLEKLTK 545
Qy 335 LMSALDGF---NII-----RYQVFTFKMNDKGEAIDKGNLVTDSKSLVLFCKDDXEYT 385
Db 546 ALKYMEDYSLRNIIVVEKELKYKNLISKIENIEIETLVE--NIKDBEQLF-----EKKIT 598
Qy 386 GEDKFNVEAIKEDGSMLEFIDTKPVLNSMDKNYFNPSKNIYVNPPEFYLRGKISDKGGF 445
Db 599 KDNKPDEKILEVSDIVKVQVQKVL---MNKIDELKKTOLILKNVE--LKHNIHVPNSY 653
Qy 446 NWEIUVNESVVDNLYIGDLHIDTRDFNIKL-----NVK----- 480
Db 654 KOENKQE-----PYYLIVLKEIDKLVFMPKPVESLINEEKKNIKTEGQSDNSEPSTEHEI 709
Qy 481 -----DGDIMDMGMKDYK-----ANGFPDKV 501
Db 710 TGOATTKPGQAGSALSGDSVQAQAEQKQAPPPVPVPEAKAQVPTPPAPVNNKTENV 769
Qy 502 TDMGNYVLQGYSDLNA-----KAVGVHYQFLYDNVKPEVNIIDPKGNTSIEYADGKSUV 556
Db 770 SKLD---YLEKLYEFLNTSVICHKYILVSHSTMNEKILQYKITKEESKLSGCCDPLDLL 826
Qy 557 FNIN-----DKRNGFDG---EIQEQHIYINGKEYTSFNDIKQIIDKTLNIKIV 603
Db 827 FNIOQNNIPVYMSFDSLNSLSQLFMEIYEKEMVCNLYKLDNDKIKNLEEAKKVSTSV 886
Qy 604 KDFARNTTVKEFIILNKDGTGEVSELKPHRVTVTIQNGKEMSSTIVSEBDFILPVYKGELEK 663
Db 887 KTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNSLKLFEIILS-----LGKNKN 937
Qy 664 GYQFDGWEISGFEGK-----KDAGVVINLSKDTFIKVPFKKIEKKBEENKPTF--D 713
Db 938 IYQ---ELIQKGSSENFYEKILKDSDTFVNESFTNFVKSADDDINSINDESKERKLEED 993
Qy 714 VSKKKDNPPQVNHNSQINSHRKEDELOREHSQKSDSTKDVATATVLDKNNISSK-STTNPP 771
Db 994 INKLKTLQLSFDLYNKYKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK 1052

Search completed: June 14, 2006, 22:52:20
Job time : 186 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:35:57 ; Search time 52 Seconds
(without alignments)
1301.176 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAESKFNKLGNGKEGS.....ATVLDKNINSSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 segs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /EMC_Celerra_SIDS3/ptodata/2/iaa/5_COMB.pcp:*
- 2: /EMC_Celerra_SIDS3/ptodata/2/iaa/6_COMB.pcp:*
- 3: /EMC_Celerra_SIDS3/ptodata/2/iaa/7_COMB.pcp:*
- 4: /EMC_Celerra_SIDS3/ptodata/2/iaa/H_COMB.pcp:*
- 5: /EMC_Celerra_SIDS3/ptodata/2/iaa/PCUS_COMB.pcp:*
- 6: /EMC_Celerra_SIDS3/ptodata/2/iaa/RE_COMB.pcp:*
- 7: /EMC_Celerra_SIDS3/ptodata/2/iaa/backfiles.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	4026	100.0	773	2 US-09-590-991-8
2	3789.5	94.1	2138	2 US-09-583-110-5274
3	3048	75.7	637	2 US-09-107-433-3169
4	750.5	18.6	1529	2 US-09-107-433-4771
5	615	15.3	117	2 US-08-961-083-68
6	615	15.3	117	2 US-09-536-784-68
7	615	15.3	117	2 US-09-765-271-68
8	615	15.3	117	2 US-09-765-272A-68
9	218	5.4	1602	2 US-09-269-874A-7
10	218	5.4	1621	2 US-09-269-874A-5
11	218	5.4	1639	2 US-09-269-874A-3
12	203	5.0	1166	2 US-09-200-650E-7
13	200.5	5.0	1786	2 US-08-973-462-8
14	191.5	4.8	2504	2 US-09-328-352-5821
15	188.5	4.7	670	2 US-09-107-433-4976
16	188.5	4.7	708	2 US-09-583-110-3019
17	187.5	4.7	1529	1 US-08-728-470-10
18	187.5	4.7	1529	2 US-08-719-641-10
19	186.5	4.6	930	2 US-09-200-650E-3
20	186	4.6	1315	2 US-09-200-650E-5
21	185	4.6	2733	2 US-09-949-016-11433
22	185	4.6	3259	2 US-09-949-016-6507
23	184	4.6	1588	5 PCT-US93-07261-11
24	184	4.6	1663	5 PCT-US93-07261-16
25	181.5	4.5	778	2 US-09-134-001C-3868
26	179.5	4.5	746	2 US-09-710-279-652

27	179.5	4.5	1183	1 US-08-447-031A-2	Sequence 2, Appli
28	179	4.4	2142	2 US-09-540-236-3459	Sequence 3459, Ap
29	178	4.4	2375	2 US-09-538-092-1131	Sequence 1131, Ap
30	177.5	4.4	1600	1 US-08-617-697-10	Sequence 10, Appl
31	177	4.4	800	7 5183745-3	Patent No. 5183745
32	174.5	4.3	1073	2 US-09-206-942-49	Sequence 49, Appl
33	174.5	4.3	1073	2 US-10-193-764-45	Sequence 45, Appl
34	174.5	4.3	1079	2 US-09-206-942-47	Sequence 47, Appl
35	174.5	4.3	1079	2 US-10-193-764-43	Sequence 43, Appl
36	174.5	4.3	10182	2 US-09-134-001C-3159	Sequence 3159, Ap
37	174	4.3	3070	2 US-09-961-403-7	Sequence 7, Appli
38	173.5	4.3	2353	2 US-09-377-155-33	Sequence 33, Appl
39	173.5	4.3	2353	2 US-08-913-942-4	Sequence 4, Appli
40	173.5	4.3	2353	2 US-09-669-974-33	Sequence 33, Appl
41	173.5	4.3	2353	2 US-09-797-862-33	Sequence 33, Appl
42	173.5	4.3	2353	2 US-09-684-707-4	Sequence 4, Appli
43	173.5	4.3	2411	2 US-09-268-347-36	Sequence 36, Appl
44	173	4.3	930	2 US-09-134-001C-5314	Sequence 5314, Ap
45	173	4.3	930	2 US-09-386-962C-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-590-991-8
; Sequence 8, Application US/09590991
; Patent No, 6887480
; GENERAL INFORMATION:
; APPLICANT: Adamou, John
; APPLICANT: Choi, Gil
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
; FILE REFERENCE: 469201-475
; CURRENT APPLICATION NUMBER: US/09/590,991
; CURRENT FILING DATE: 2000-06-09
; EARLIER APPLICATION NUMBER: U.S. 60/138,453
; EARLIER FILING DATE: 1999-06-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 773
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-590-991-8

Query Match	100.0.0%;	Score	4026;	DB 2;	Length	773;	
Best Local Similarity	100.0.0%;	Pred. No.	3.6e-254;				
Matches	773;	Conservative	0;	Mismatches	0;	Gaps	0;
Qy	1	KLGEIAESKFNKLGNGKEGSLKDTTGVVEHHHGHNEESIKEKSSFTIDRNISTIRDFENK	60				
Db	1	KLGEIAESKFNKLGNGKEGSLKDTTGVVEHHHGHNEESIKEKSSFTIDRNISTIRDFENK	60				
Qy	61	DLKKLKKKPREVDFTSETGKMEEEYDYKYDDKGNIIAYDDGTDLEYTEKELDEISKI	120				
Db	61	DLKKLKKKPREVDFTSETGKMEEEYDYKYDDKGNIIAYDDGTDLEYTEKELDEISKI	120				
Qy	121	YGVLSPSKDGHPFELGKISNVSNKAKVYGNNYKSIIEIKATKYDFHKTMTFDFLYANIND	180				
Db	121	YGVLSPSKDGHPFELGKISNVSNKAKVYGNNYKSIIEIKATKYDFHKTMTFDFLYANIND	180				
Qy	181	IVDGLAPAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN	240				
Db	181	IVDGLAPAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN	240				
Qy	241	KPDNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVTTYNPFGKTMLESGNGVYSKEDI	300				
Db	241	KPDNLTKMESGKIYSDSEKQYLLKDNIIILRKGVALKVTTYNPFGKTMLESGNGVYSKEDI	300				
Qy	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKMDKGEA	360				
Db	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKMDKGEA	360				

Qy 361 IDKGNLVTSSKLVLFKGGKDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 361 IDKGNLVTSSKLVLFKGGKDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Qy 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHTDNTDRDNFKLVNK 480
Db 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHTDNTDRDNFKLVNK 480
Qy 481 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFLYDNVKNPEVNID 540
Db 481 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFLYDNVKNPEVNID 540
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGPDGEIQBOHIIYINGKEYTSFNDIKQIIDKTLNLIK 600
Db 541 PKGNTSIEYADGKSVVFNINDKRNNGPDGEIQBOHIIYINGKEYTSFNDIKQIIDKTLNLIK 600
Qy 601 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIIVSEEDFILPVYKGE 660
Db 601 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIIVSEEDFILPVYKGE 660
Qy 661 LEKGYQFDGWEISGFEGKKGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKKDN 720
Db 661 LEKGYQFDGWEISGFEGKKGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKKDN 720
Qy 721 PQVHSQLNESHKEDLQREHSHQSDSTKDVDTATVLDKNNISSKSTNNPNK 773
Db 721 PQVHSQLNESHKEDLQREHSHQSDSTKDVDTATVLDKNNISSKSTNNPNK 773

RESULT 2
US-09-583-110-5274
; Sequence 5274, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT FILING DATE: 2000-05-26
; PRIOR FILING DATE: 1997-07-02
; PRIOR FILING DATE: 1998-06-30
; PRIOR FILING DATE: 1998-05-12
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5274
; LENGTH: 2138
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5274

Query Match 94.1%; Score 3789.5; DB 2; Length 2138;
Best Local Similarity 94.8%; Pred. No. 3.7e-238;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;
Qy 1 KLGIATSKFNKLGKESGLKDDTTGVEHHQHNESIEKSSFTIDRNIETIRDFENK 60
Db 1333 KLGIATSKFNKLGKESGLKDDTTGVEHHQHNESIEKSSFTIDRNIETIRDFENK 1392
Qy 61 DLKLLIKKKFREVDDFTSETGRMEVDYKDDKGNIIAYDDGDTLLEYETKLEIKSKI 120
Db 1393 DLKLLIKKKFREVDDFTSETGRMEVDYKDDKGNIIAYDDGDTLLEYETKLEIKSKI 1451
Qy 121 YGVLSPSKDGHEILGKISNKNKAYVGNVYKSIETKATYDFHSTMTFDLYANIND 180
Db 1452 YGVLSPSKDGHEILGKISNKNKAYVGNVYKSIETKATYDFHSTMTFDLYANIND 1511
Qy 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIBELGSDLSKN 240
Db 1512 IVDGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIBELGSDLSKN 1571

Qy 241 KPNLNTMESGKIYSSEKQOYLLKUNIIILRGYALKVTTYNFGKTDMLBNGVYISKEDI 300
Db 1572 KPNLNTMESGKIYSSEKQOYLLKUNIIILRGYALKVTTYNFGKTDMLBNGVYISKEDI 1631
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFKQNDKGEA 360
Db 1632 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIIRYQVTFKQNDKGEA 1691
Qy 361 IDKGNLVTSSKLVLFKGGKDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1692 IDKGNLVTSSKLVLFKGGKDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 1751
Qy 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHTDNTDRDNFKLVNK 480
Db 1752 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHTDNTDRDNFKLVNK 1811
Qy 481 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFLYDNVKNPEVNID 540
Db 1812 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLTQGYSDLNKAKAVGVHYQFLYDNVKNPEVNID 1871
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGPDGEIQBOHIIYINGKEYTSFNDIKQIIDKTLNLIK 600
Db 1872 PKGNTSIEYADGKSVVFNINDKRNNGPDGEIQBOHIIYINGKEYTSFNDIKQIIDKTLNLIK 1931
Qy 601 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIIVSEEDFILPVYKGE 660
Db 1932 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIIVSEEDFILPVYKGE 1991
Qy 661 LEKGYQFDGWEISGFEGKKGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKKDN 720
Db 1992 LEKGYQFDGWEISGFEGKKGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKKDN 2051
Qy 721 PQVHSQLNESHKEDLQREHSHQSDSTKDVDTATVLDKNNISSKSTNNPNK 773
Db 2052 PQVHSQLNESHKEDLQREHSHQSDSTKDVDTATVLDKNNISSKSTNNPNK 2104

RESULT 3
US-09-107-433-3169
; Sequence 3169, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3169:
SEQUENCE CHARACTERISTICS:
LENGTH: 637 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...637
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169
Query Match 75.7%; Score 3048; DB 2; Length 637;
Best Local Similarity 99.3%; Pred. No. 1.5e-190;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 185 LAFAGDMRLFVKNDQKAEIKIRPEKIKETKSEYVYSSYGNVIELGEGDLSKNKPDN 244
Db
QY 245 LTKMESGKIYSSEKQYLLKNIILIRKGYALKVTTYPNGKTDMLGNGVYSKEDIKIQ 304
Db
QY 75 LTKMESGKIYSSEKQYLLKNIILIRKGYALKVTTYPNGKTDMLGNGVYSKEDIKIQ 134
QY 305 KANPNLRALSETTIYADSRNVEDGSTQVLSALDGNIIIRYQVTFPMNDKGAIDKD 364
Db
QY 135 KANPNLRALSETTIYADSRNVEDGSTQVLSALDGNIIIRYQVTFPMNDKGAIDKD 194
QY 365 GNLVTDSSKLVLFSGKDDKEYTGEDKFNVEAIKEDGSMFLFIDPKPNLSMDKNYFNPSKSN 424
Db
QY 195 GNLVTDSSKLVLFSGKDDKEYTGEDKFNVEAIKEDGSMFLFIDPKPNLSMDKNYFNPSKSN 254
QY 425 KIYVRNPFYLRGKISDKGFNWLVRNESVVDNLYIGDLHIDNTRDFNIKLVKGDGI 484
Db
QY 255 KIYVRNPFYLRGKISDKGFNWLVRNESVVDNLYIGDLHIDNTRDFNIKLVKGDGI 314
QY 485 MDWGMKDYKANGFPDKVTMDGNVLOTGYSDLNAKAVGVHVFQYLDNVKPEVNDPKGN 544
Db
QY 315 MDWGMKDYKANGFPDKVTMDGNVLOTGYSDLNAKAVGVHVFQYLDNVKPEVNDPKGN 374
QY 545 TSIEYADGKSVVFNINDKRNNGFDEIOEQHIYINGKEYTSFNDIKIIDKTLNKKIVVK 604
Db
QY 375 TSIEYADGKSVVFNINDKRNNGFDEIOEQHIYINGKEYTSFNDIKIIDKTLNKKIVVK 434
QY 605 DPARNTTVKGFILNKDGTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKG 664
Db
QY 435 DPARNTTVKGFILNKDGTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKG 494
QY 665 YQFDGWEISGPEGKDGAVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKKNPOVN 724
Db
QY 495 YQFDGWEISGPEGKDGAVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKKNPOVN 554
QY 725 HSQLNESHKEDLQREHSQKSDSKDTATVLDKNNISSKSTNNPNK 773
Db
QY 555 HSQLNESHKEDLQREHSQKSDSKDTATVLDKNNISSKSTNNPNK 603
RESULT 4
US-09-107-433-4771
; Sequence 4771, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4771:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1529
SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-09-107-433-4771
Query Match 18.6%; Score 750.5; DB 2; Length 1529;
Best Local Similarity 79.7%; Pred. No. 2.6e-40;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;
QY 1 KLGEIAESKFNKLGKSGSLKDDTTGVVHHQHEESIKESKSFITDRTSTIRDFENK 60
Db 1336 KLGEIPESKFNKLGKSGSLKDDTTGVVHHQHEESIKESKSFITDRTSTIRDFENK 1395
QY 61 DLKKLKKKFFREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 120
Db 1396 DLKKLKKKFFREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 1454
QY 121 YGVLSKSGKHGFEILGKISNVSKNAKYVGNVYNNYSIEIKATKYPHSTMTFDPYANIND 180
Db 1455 YGVLSKSGKHGFEILGKISNVSKNAKYVGNVYNNYSIEIKATKYPHSTMTFDPYANIND 1514
QY 181 IVDGLAF 187
Db 1515 IVDGLAF 1521
RESULT 5
US-08-961-083-68
; Sequence 68, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville

```
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Dikette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-68

Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 657 YKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKIEKKKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKIEKKKEENKPTFDVSK 60

Qy 717 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 117

RESULT 6
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Dikette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961.083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
```

```
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 657 YKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKIEKKKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKDGAGYVNLKDTFIKPVFKIEKKKEENKPTFDVSK 60

Qy 717 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 117

RESULT 7
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Dikette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/535,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-271-68
```

```
Query Match      15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YGELEKGQFDGWISGEGKDGAGYVNLKSDTFIKPVFKIEBKKEENKPTFDVSK 716
Db 1 YGELEKGQFDGWISGEGKDGAGYVNLKSDTFIKPVFKIEBKKEENKPTFDVSK 60

QY 717 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 773
Db 61 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117

RESULT 8
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
; Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: PB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68
```

```
Query Match      15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YGELEKGQFDGWISGEGKDGAGYVNLKSDTFIKPVFKIEBKKEENKPTFDVSK 716
Db 1 YGELEKGQFDGWISGEGKDGAGYVNLKSDTFIKPVFKIEBKKEENKPTFDVSK 60

QY 717 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 773
Db 61 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117
```

RESULT 9

```
US-09-269-874A-7
; Sequence 7, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; FILE OF INVENTION: Complete Malaria Antigen, GPI90/MSPI
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1602
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-09-269-874A-7
```

```
Query Match      5.4%; Score 218; DB 2; Length 1602;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

QY 23 KDTTG-VEHHQNEESIK-----EKSSFTIDRNIISTIRDFENKDLKLIKKEFEVDD 75
Db 213 KDNVGMEDYIKNNKKTIENTINELIEESKKTIDKNKNATKEE-----KKKLYQAOY 264

QY 76 FTSETGKMEEEYDYKDDGNIIAYDDGTLEYTEKLEDEIKSIKIVLSPSKDGHFEL 135
Db 265 DLSIYNKQLEE-----AHNLISV-----LE---KRIDLK-----KNEIKELL 300

QY 136 KGISNVSNKAVYGNVNNYKSIKATKYDFH-----SKMTFOLYANINDIVDLAFA 188
Db 301 DKINEIKNPPPPANSNGTPTLLDKNKKTEEHEKEIKETAKTIKFNIDSIFTDPL----- 354

QY 189 GDMRLFVKDNDOKKAEIKIRMEKIKETKSEVP-----VSSYXGNV-----IELGE-GDLSK 239
Db 355 -ELEYLREKN-KNIDISAKVETKESTPNFNGVTYPLSYNDINNALNELNSFGDLI- 411

QY 240 NKPDNLTKMESGKIYSDSEKQOYL--LKDNIL-----RKGYALKVTTYNPGKTDML 289
Db 412 -NPFDTKEPSKNIYTDNERKKFNEIKEIKIEKKKTESDKSVEDSKSLN-----DIT 466

QY 290 EG-----NGVYSKE-----DIAKTOKA-----NPNLRALSETTIYVADSRNVEDGRSTOSV 334
Db 467 KEYEKLNLNEIYDSKFNNNIDLTNFEKMMGKRYSVKVEKLTHTHTNTPASVENSKHNLKLT 526

QY 335 LMSALDGF--NII-----RYQVFTFMNDKGEAIDKGNLVTDSKLVLFKDDKEYT 385
Db 527 ALKYMEDYSLRNI VVEKELKYKNLISKIENIEITLVE--NIKKDEEQLF-----EKKIT 579

QY 386 GEDKFNVEAIAKEDGSMFLFDTKPVNLNMDKNVFNPSKNIYVVRNPFYLRGKISDKGKF 445
Db 580 KDNKPKDEKILEVSDIVKQVQKVL---MKNIDELKKTQILKQV---LKHNIHVPSY 634

QY 446 NWELRVNESVVDNYLIYGDHLHDNTRDPNKL-----NVK-----480
Db 635 KQENKQE-----PYVLIVLKKEIDKLKVPMPKVESLINEEKKNIKTEGQSDNSEPSTEGEI 690

QY 481 -----DGDINDWGMKDYK-----ANGFPDKV 501
Db 691 TQQTATTPGQAGSALGDSVQAQEQOQAPVPVPVPEAKAQVPTTPAPVNNKTNV 750

QY 502 TDMGNYVLOTGYSDLNA-----KAVGVHYOFLYDYNKPEVNIIDPKGNTSEYADGKSV 556
Db 751 SKLD---YLEKLYEFLNTSYICHKYLIVSHSTMNEKILKQYKIKTEBESKSSCDPLDL 807

QY 557 FNIN-----DKRNNPGDG---EIQEOHIYINGKEYTSFNDIKOIIDKTLNIKIV 603
Db 808 FNIQNNIPVMYSMPFDSLNNLSQLFMIEIYKEMVNCNLYKLKDNNDIKNLLBEAKVSTSV 867
```

```
Qy 604 KDFARNTTVKEFILNKDTGEVSELKPHRVTVTQNGKEMSSSTIVSEEDFILPVYKGELEK 663
Db 868 KTLSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFENILS-----LGNKN 918
Qy 664 GYQFDGWEISGEFK-----KDGAVINLSKDTPIKVPFKKIBKKKEENKPTF--D 713
Db 919 IYQ-----ELIGQKSENFEYKILKDSDTFYNESFTNFKVKADDINSLNDESKRKKLEED 974
Qy 714 VSKKDNQPNVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSK-STTNNP 771
Db 975 INKLKKTQLSFDLYNKYKLERLFDKPKKTGVYKQKIKLTLKLEQLESKLSLNNP 1033

RESULT 10
US-09-269-874A-5
; Sequence 5, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSp1
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1621
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-5

Query Match 5.4%; Score 218; DB 2; Length 1621;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VEHHHQENESIK-----EKSFTIDRNISTIRDFENKDLKLIKFKFREVDD 75
Db 232 KDNVGMEDYIKKKNKTNIENILIESKKTIDKNKNATKEE-----KKLYQAOY 283
Qy 76 FTSETGKRMEYDYKYDDKGNIIAYDDGTDLEYETEKLEIKSIYGLVSPSKDGHFEIL 135
Db 284 DLSIYNKQLEE-----AHLNISV-----LE---KRIDLK-----KNENIKELL 319
Qy 136 GKISNVSKNAKVYGNKYSEIKATKYDPH-----SKTMTFDLYANINDIVDGLAFA 188
Db 320 DKINEIKNPPANGSGTNPNTLLDKNKKIEHEKEIKIAKTIFKNIDSLFTDPL----- 373
Qy 189 GDMRLFVKNDQKAEIKIRMPKIKETKSEYP-----YVSSYGNV-----TELGE-GDLSK 239
Db 374 -ELEYLRKN-KNIDISAKVETKESTPNPNGVTYPLSYNDINNALNELNSFGDLI- 430
Qy 240 NKPNLTLMESGKIYDSEKQOYL--LKDNIL-----RKYALKVTTYPNGKTDML 289
Db 431 -NPPDYTKPSKNITVDNERKKFNEIKIEKIEKIESDKSYEDRSKSLN---DIT 485
Qy 290 EG-----NGVYSKE-----DTAKTQKA-----NPNLRSLETTIYASRVNVEDGRSTQSV 334
Db 486 KEYEKLNLNIYDSKFNNDIDTNFKNMGKRYKVEKLTHHNTFASYENSKHLEKLTK 545
Qy 335 LMSALDGF--NII-----RYQVTFKQNDKGBAIDKGNLVTDSSKLVFGKDDKEYT 385
Db 546 ALKTMEDYSLRNIIVVEKELKYKNLSKIENIEITLVE--NKKDEQLF-----EKKIT 598
Qy 386 GEDKFNVBALIKEDGSMFLFDTKPVLNSMDKNYFNPSKSNKIYVRNPFYLGKISDKGGF 445
Db 599 KDNKPKDEKILEVSDIVKQVQKVL---MNKIDELKKTQLILKNVE--LKHNIHVNSY 653
```

```
Qy 446 NMELRVNESVVDNYLIYGDHLHDNTDFFNKL-----NVK----- 480
Db 654 KOENKQE-----PYLLIVLKKEDIKLKFVMPKVESLINEKKNIKTSGQSDNSPSTGEI 709
Qy 481 -----DGDIMDMGME-DYK-----ANGFPDKV 501
Db 710 TGOATTTKPGQAGSALEGDSVQAQAEQKQAQPPVPVPVPEAKAQPVTTPPAPVNNKTENV 769
Qy 502 TMDGNNVYQTYGSD,NA-----KAVGVHYQFYLDNVKPEVNDPKNGTISIEYADGKSVV 556
Db 770 SKLD---YLEKLYEFLNTSYICHKYLIVSHSTNNEKILKQYKITKEESKLSKCDPLDLL 826
Qy 557 FNIN-----DKRNNGFDG---EIQEOHIYINGKEYTSENDIKIIDLKTLNLIKIV 603
Db 827 FNIONNIPVMYSMFDLSLNSLSOLFMEIYEKEMVCNLYKLDKNDKIKNLLLEEAKKYSTSV 886
Qy 604 KDFARNTTVKEFILNKDTGEVSELKPHRVTVTQNGKEMSSSTIVSEEDFILPVYKGELEK 663
Db 887 KTLSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFENILS-----LGNKN 937
Qy 664 GYQFDGWEISGEFK-----KDGAVINLSKDTPIKVPFKKIBKKKEENKPTF--D 713
Db 938 IYQ-----ELIGQKSENFEYKILKDSDTFYNESFTNFKVKADDINSLNDESKRKKLEED 993
Qy 714 VSKKDNQPNVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSK-STTNNP 771
Db 994 INKLKKTQLSFDLYNKYKLERLFDKPKKTGVYKQKIKLTLKLEQLESKLSLNNP 1052

RESULT 11
US-09-269-874A-3
; Sequence 3, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSp1
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-3

Query Match 5.4%; Score 218; DB 2; Length 1639;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VEHHHQENESIK-----EKSFTIDRNISTIRDFENKDLKLIKFKFREVDD 75
Db 232 KDNVGMEDYIKKKNKTNIENILIESKKTIDKNKNATKEE-----KKLYQAOY 283
Qy 76 FTSETGKRMEYDYKYDDKGNIIAYDDGTDLEYETEKLEIKSIYGLVSPSKDGHFEIL 135
Db 284 DLSIYNKQLEE-----AHLNISV-----LE---KRIDLK-----KNENIKELL 319
Qy 136 GKISNVSKNAKVYGNKYSEIKATKYDPH-----SKTMTFDLYANINDIVDGLAFA 188
Db 320 DKINEIKNPPANGSGTNPNTLLDKNKKIEHEKEIKIAKTIFKNIDSLFTDPL----- 373
Qy 189 GDMRLFVKNDQKAEIKIRMPKIKETKSEYP-----YVSSYGNV-----TELGE-GDLSK 239
Db 374 -ELEYLRKN-KNIDISAKVETKESTPNPNGVTYPLSYNDINNALNELNSFGDLI- 430
Qy 240 NKPNLTLMESGKIYDSEKQOYL--LKDNIL-----RKYALKVTTYPNGKTDML 289
```

Db 431 -NPFDTKPSKNIYTDNERKFFINEIKIEKKIEBSDKSYEDRSKSLN----DIT 485
Qy 290 EG-----NGVYSKE-----DIKIOKA-----NPNLRALSETTIYADSRNVEDGRSTQSV 334
Db 486 KEYEKLNLNEYDSKFNNIDLTFEKMOMGRYSYKVEKLTHTHTTFASYENSXHNLEKLTJK 545
Qy 335 LMSALDGF---NII-----RYQVTFKMDKGEAIDKDNLVTDSSKLVLFGKDKKEYT 385
Db 546 ALKYMEDYSLRNIVVEKELKYKLNLSKIENBETLVE--NIKKDEQLF-----EKKIT 598
Qy 386 GEDKFNVEAIKEDGSMFLFDTKFPVNSMDKNVFNPSKSNKIYVRNPFYLRKISDKGKF 445
Db 599 KOENKPEKILEVSDIVKVQVQKVL--MNIKIDELKTLQLKKNVE--LKHNIHVPNSY 653
Qy 446 NWELRVNESVDNLYLGLHDINTREDNKL-----NVK----- 480
Db 654 KOENKQE-----PYLLVLKKEIDKLKVFMPKVESLINEEKKNIKTEGQSDNSEPSTGEI 709
Qy 481 -----DGDIMDMGMKDYK-----ANGFPDKV 501
Db 710 TGOATTKPGQAGSALGSDSVQAQAQEQQAQPPVPVPEAKAQVPTPPAPVNNKTENV 769
Qy 502 TMDGNVYLQTVSDINA-----KAVGVHYQFLYNVKNVPEVNDPKNTSIBYADGKSVV 556
Db 770 SKLD---YLEKLYEFLNTSYICHKYLIVSHSTWNEKILKQYKITKEESKLSGCDPLDL 826
Qy 557 FNIN-----DKRNNFGD---EIOEQHIYINGKEYTSFNDIKQIIDTLNLIKIVV 603
Db 827 FNIOINIPWYMFSLNLSLSQLFMEIYEKEMVNCNLYKLKDNKIKNLLEAKKVSTSV 886
Qy 604 KDFARNTTVKEFILNKDTEVSELKPHRVTVTIQNGKEMSVTIVSEEDFILPVYGELEK 663
Db 887 KTLSSSSMQPLSITPDQKEVGSANDDTSHSTWNLNLSKLFPENLS-----LQKXN 937
Qy 664 GYQFGWEISGPEGK-----KQAGYVINISKDTFFIKVPKPIEEKKEENKPTF--D 713
Db 938 IQYQ---ELIGQKSENFEKILKSDTFYNESFTNFVKSADDINSLNDESKRKKLEED 993
Qy 714 VSKKXNDPWNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNISSK-STNNP 771
Db 994 INKLKKTQLSPDLNKKYKLERLDFDKKTVGKYMQIKKLTLLKEQLSKLSLNLP 1052

RESULT 12

US-09-200-650E-7
; Sequence 7, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Eihinn, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-7

Query Match 5.0%; Score 203; DB 2; Length 1166;
Best Local Similarity 21.6%; Pred. No. 8.7e-05;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

Qy 33 QENESSIKE--KSSFTIDENISTIRDFENKDKLKKLKKFR-----EYDDFTS 78
Db 226 KNNPEKLKELVNDNNTDHTKPVATAPTSVAPKRVNAKMFAPAFAVAAVANNVNDLTK 285
Qy 79 ETGKEMEEDYKYDDKGNIIAYDDGTDLLEYETE-----KLDIEIKSKIYGLVS 125
Db 286 VTKQITIK-----VGDKDNVAAAHDKDIEYDTEFTIDNKVKKGDWTINVDKNVPSDLT 341
Qy 126 PSKD-----GHEIILGKISNVSKNAK---VYGGNNYKSIIEIKATKYDYFHSK----- 168
Db 342 DKNDPIDITDPSGEVIAGTFDKATKQITPTFDYVDKYEDIKSRLLTLYSYIDKKKTPVNE 401
Qy 169 ---TWTF-----DLVANIN-DIVDGLAFAGD---NRLFVK-DNDQKKAIEKIRM-PEKIK 214
Db 402 TSLNUTPATAGKETSONVTVDYQDPWVH-GDSNIOISIFTKLDEDKQTTEQQIYVNPPLKKS 460
Qy 215 ETKSY-----PYVSSYGNVIELGED--LSKNKPNLTKM-----ESGKIYSDEKQ 261
Db 461 ATNTKVDIAGSQVDDYGN-IKLNGSTIIDQNTETIKVYKVNSDQQLPQSNRIYDFSOYED 519
Qy 262 YLLK-DNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDIKIQANPNLRALSETTIYA 320
Db 520 VTSQFDN---KKSFSNNVATLDFG-----DINSAYI-----IKVSVKTYPTS 558
Qy 321 DSR-NVEDGRSTQSVLMSALDGFNIIRYOVTFFKMDKG-----E 359
Db 559 DGEELDIAOGTSNRIT--DKYGYNYAGYSNFIIVTSNDTGGGDTVKPBEKLYKIGDYVWE 616
Qy 360 AIDKGNLVTDSSKLVLFGKDDKEYTGRDKFNVEAIKEDGSMFLFDTKPVNLSMDKNY-F 418
Db 617 DVDRKG-----VQGTDSKEKPMANVL-VLTYPDGT-----TKSVRTDANGHVEF 660
Qy 419 NPSKSNKIYV---RNPEFYLRGKIS-----DKGGFNWELRVNVSVDNLYLIGDLHD 468
Db 661 GGLKDGETYTVKFETPTGYLPTKVGTTDGEKDSNGSSVTVKINGK-----DDMSLD 712
Qy 469 -----NTRDFNIKLVNKG--DINDGMKQYKANGFPDKVTDMDGNVYLQYSD 517
Db 713 TGFYKPEKYNLGDYVWEDTNKDGIDQANEPGDKVKVT-----LKDSTGKV-IGTTTDA 766
Qy 518 NAKAVGVHYQFLYDNVKNPEVNDPKNTSIEY---ADGKSVVFN--INDKRNNGFDGEIQ 572
Db 767 SGK-----YKF-----TDLD-NGNVTVEFEIPAGYTPVKNTTADDKOSNGL----- 807
Qy 573 EQHIYINGKEYTSFNDIKQIIDKTLNLIKIVVKDFARNT--TVKEFI---LNKDTGEVSEL 627
Db 808 -----TTTGVIKADADNMTLTLD-----RGFYKPKYSLGIVVYVYDSNKGKQDSTE 851
Qy 628 KPHR-VTVTIQNGK-EMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKQDAGYVI 685
Db 852 KGIKDVTVTLQNEKEVIGITKTIDEN-----GK-----YRFD----- 883
Qy 686 NLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDNPPQV-----NHSQ--LNESHKEDLQ 738
Db 884 NLDGSKY-KVIFEXKPAIGTQVTWTTED-DKADGGEVDVITDHDHDFLNGYFEEDTS 941
Qy 739 REHSQKSDSTKD 751
Db 942 DSDSDSDSDSDSD 954

RESULT 13

US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06

THIS PAGE BLANK (uapfo)